DOCUMENT RESUME

ED 356 668 FL 021 178

AUTHOR Hartwell, Tyler; And Others
TITLE Design, Data Collection and Analysis of a Field Test of Instruments and Procedures To Measure English Language Proficiency.
INSTITUTION Research Triangle Inst., Research Triangle Park, N.C.
SPONS AGENCY National Center for Education Statistics (DHEW), Washington, D.C.
PUB DATE Apr 76
CONTRACT 300-75-0253
NOTE 150p.; For a related document, see FL 021 179.
PUB TYPE Reports - Descriptive (141) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Bilingualism; Data Collection; English (Second Language); *Language Proficiency; *Language Tests; *Limited English Speaking; *Measurement Techniques; National Surveys; *Questioning Techniques; Research Methodology; Test Construction; Test Validity

ABSTRACT

A project to field test instruments and procedures for measuring English language proficiency, to be used in a large-scale national household survey, is reported. The major purposes of the field test were to measure the validity of census-type questions for determining language proficiency, using several different bilingual ethnic groups, and to evaluate proposed data collection procedures for the larger study. The report begins with an overview of the project. Chapter 2 explains the design of the original field test, revisions made to it, and site selections. The third chapter describes sampling activities: securing school district cooperation, obtaining lists of students, and selection of sample individuals. Samples were drawn from Miami (Florida), El Paso (Texas), San Francisco (California), and Ganado and Window Rock (Arizona). Data collection activities are outlined in chapter 4, and analysis activities are described and results are presented in chapter 5. Appended materials include the bilingual study census questionnaire and compilations of additional results of discriminate analysis. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED356668

RESEARCH TRIANGLE INSTITUTE

RTI Project No. 250U-1148

April, 1976

DESIGN, DATA COLLECTION AND ANALYSIS OF A FIELD TEST OF INSTRUMENTS AND PROCEDURES TO MEASURE ENGLISH LANGUAGE PROFICIENCY

by

Tyler Hartwell
Paul Moore
Michael Weeks
Robert Mason
and
Babu Shah

Research Triangle Institute
Research Triangle Park, North Carolina 27709

CAL Project Director: Walter S. Stolz

Prepared for

Center for Applied Linguistics
1611 North Kent Street
Arlington, Virginia 22209

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Prime Contract No. 300-75-0253

BEST COPY AVAILABLE

FL 021 178

ACKNOWLEDGEMENTS

The authors would like to acknowledge the valuable input provided by Dr. Walter Stolz and Mrs. Maggie Bruck of the Center for Applied Linguistics and Mr. Leslie Silverman and Mrs. Vickie Kojcsich of the National Center for Educational Statistics in carrying out the research described in this report.

In addition, we would like to thank Mrs. Jean Allen, Mrs. Brenda Gurley and Ms. Deborah Whitehurst for their untiring efforts in preparing the final report.

TABLE OF CONTENTS

		Page
	LIST OF TABLES	iii
I.	INTRODUCTION	1
II.	DESIGN OF FIELD TEST	2
	A. Overview	2
	B. Original Proposed Design	2
	C. Revised Design	4
	D. Selection of Sites	5
III.	SAMPLING ACTIVITIES	7
	A. Securing School District Cooperation	7
	B. Obtaining Lists of Students	8
	C. Selection of Sample Individuals	8
IV.	DATA COLLECTION ACTIVITIES	12
	A. Overview	12
	B. Design of the Data Collection Plan	13
	C. Instrument Development	15
	D. Recruitment of Interviewers	18
	E. Final Preparations for Training and Data Collection	21
	F. Training of the Field Staff	22
	G. Data Collection Procedures	23
	H. Fieldwork Results	28
	I. Processing Completed Field Test Instruments	31
	J. Confidentiality Procedures	32
	K. Attendance at Language Group Representative (LGR) Meetings	32
V.	ANALYSIS ACTIVITIES	34
	A. Overview	34
	B. Description of Sample Members	36
	C. Identifying Useful Census Questions	47
	D. Factor Analyses	52
	E. Regression Analyses	55
	F. Discriminant Analyses	70
	F.1 Discriminant Functions for Children	71
	F.2 Overall Discriminate Function	79
	F.3 A Discrete Discriminating Rule	82
	F.4 Estimation of Percent LESA	86
	F.5 Discriminant Functions for Adults	97
	G. Reverse Scale Questions	108
	H. Monolingual - Bilingual Interviewer Comparisons	113

Page

REFERENCES 116

APPENDIX A: Bilingual Study Census Questionnaire

APPENDIX B: Additional Results of Discriminate Analysis Using FCTR and CP to Define LESA

LIST OF TABLES

		Page
Table III.1	Sample Sizes by Site, Age Group, and List Code Strata	10
Table IV.1	Schedule of Principal Data Collection Activities	14
Table IV.2	Data on Recruiting Interviewers for the Field Test	20
Table IV.3	Data Collection Results of Field Test	29
Table IV.4	Comparison of Data Collection Results for Monolingual and Bilingual Interviewers	30
Table V.1	Size of Household, Percentage Distribution from Item 1 of Household Information Form	37
Table V.2	Availability of Telephone, Percentage Distribution from Item 5 of Household Information Form	38
Table V.3	Education of Head of Household, Percentage Distribution from Item 6 of Household Information Form	39
Table V.4	Employment Status of Head of Household, Percentage Distribution from Item 7 of Household Information Form	40
Table V.5	Occupation of Head of Household, Percentage Distribution from Item 8 of Household Information Form	41
Table V.6	Family Income, Percentage Distribution from Item 9 of Household Information Form	42
Table V.7	Origin or Descent of Sample Children, Percentage Distribution from Item 2 of Census Questionnaire (Second Hand Reports)	43
Table V.8	Origin or Descent of Sample Adults, Percentage Distribution from Item 2 of Census Questionnaire (First Hand Reports)	44

		Page
Table V.9	State, U.S. Territory, or Foreign Country of Birth for Sample Children, Percentage Distribution from Item 3 of Census Questionnaire (Second Hand Reports)	45
Table V.10	State, U.S. Territory, or Foreign Country of Birth for Sample Adults, Percentage Distribution from Item 3 of Census Questionnaire (First Hand Reports)	46
Table V.11	List of Census Questions Satisfying Screening Criteria.	51
Table V.12	Definitions of Recoded Variables and Sites Used in Multiple Regression and Discriminate Analyses	57
Table V.13	Summary of Regression Analyses for Children With 10 Census Questions as Independent Variables and Several Different Dependent Variables, Body of Table Gives Standardized Regression Coefficients, Regressions Run by Sites and Over Sites.	63
Table V.14	Summary of Regression Analyses for Adults With 11 Census Questions as Independent Variables and Several Different Dependent Variables, Body of Table Gives Standardized Regression Coefficients, Regressions Run by Site and Over Sites.	65
Table V.15	Summary of Regression Analyses for Children With Independent Variables Spund, Usage and Years and Several Different Dependent Variables, Body of Table Gives Standardized Regression Coefficients, Regressions Run by Sites and Over Sites.	68
Table V.16	Results of Using Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on 10 Census Questions, LESA Defined by LSTCD and FCTR	74
Table V.17	Results of Using Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, LESA Defined by LSTCD and FCTR	76
Table V.18	Standardized and Unstandardized Discriminant Function Coefficients for Children for Spund, Usage and Years, LESA Group Defined by LSTCD and FCTR	78

		Page
Table V.19	Results of Using a Discriminant Function Estimated Over All Sites to Predict Whether a Child is LESA or Non-LESA Within Each Site, Discriminant Function Based on Spund, Usage and Years, LESA Defined by LSTCD and FCTR	81
Table V.20	Illustration of Cochran's Classification Procedure for the Children Field Test Data for Two Different Estimates of π_1, π_2	85
Table V.21	Comparison of Prediction Results Over Sites Using Different Classification Procedures to Predict Whether a Child is LESA or Non-LESA; Procedure Based on Spund, Usage and Years; LESA Defined by LSTCD and FCTR	87
Table V.22	Schematic of Misclassification Errors	88
Table V.23	Results of Using Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, Field Data Adjusted so that Over Sites % LESA = 26%, % Non-LESA = 74% as Defined by LSTCD	90
Table V.24	Standardized and Unstandardized Discriminant Function Coefficients for Children for Spund, Usage and Years, Field Test Data Adjusted so that Over Sites % LESA = 26%, % Non-LESA = 74% as Defined by LSTCD	91
Table V.25	Results of Using Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, Field Test Data Adjusted so that Over Sites % LESA = 39%, % Non-LESA = 61% as Defined by FCTR	92
Table V.26	Results of Using Discriminant Analysis to Predict Whether an Adult is LESA or Non-LESA, Discriminant Function Based on 11 Census Questions, LESA Group Defined by LSTCD and FCTR	101
Table V.27	Results of Using Discriminant Analysis to Predict Whether an Adult is LESA or Non-LESA, Discriminant Function Based on 6 Census Questions, LESA Group Defined by LSTCD and FCTR	103
Table V.28	Results of Using a Discriminant Function Estimated Over All Sites to Predict Whether an Adult is LESA or Non-LESA Within Each Site, Discriminant Function Based on 11 Census Questions, LESA Defined by FCTR	105

		Page
Table V.29	Cochran's Classification Procedure for the Adult Field Test Data Over Sites	106
Table V.30	Results of Using Discriminant Analysis to Predict Whether an Adult is LESA or Non-LESA, Discriminant Function Based on 11 Census Questions, Field Data Adjusted so that Over Sites % LESA = 19%, % Non-LESA = 81% as Defined by FCTR	107
Table V.31	Results of Reverse Scale Analysis for Four Census Questions for Children and Adults by Site, Body of Table Gives the Distribution of Respondents for the Normal and Reversed Scales	111
Table V.32	Sample Means and Summary of t-Tests on Monolingual Versus Bilingual Interviewer Means for Various CQs and CP Score, Data for Children for Paired Interviewers Only	115
Table B-1	Results of Using Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on 10 Census Questions, LESA Group Defined by FCTR	
Table B-2	Results of Usng Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, LESA Defined by FCTR and CP	

I. INTRODUCTION

The Research Triangle Institute (RTI), acting as a subcontractor to the Center for Applied Linguistics (CAL), conducted a field test of instruments and procedures designed to measure English language proficiency (MELP). The MELP study will be conducted in 1976 by the Bureau of the Census in conjunction with a large national household survey. The major purposes of the field test were: (1) to measure the validity of using Census-type questions to measure the English language proficiency of samples of several different bilingual ethnic groups, and (2) to evaluate proposed data collection procedures for use by the Census in conducting the MELP study.

This report describes and documents the tasks performed by RTI during the project, which began in early June of 1975 and ended in April 1976. Since the study design was modified somewhat after the project was begun, Chapter II of this report includes descriptions of both the design originally proposed and the modified design. Chapter III describes sample selection activities completed by RTI; Chapters IV and V document the data collection and analysis tasks RTI completed under the subcontract with CAL.

II. DESIGN OF FIELD TEST

A. Overview

In this chapter the original proposed design, the revised design and the selection of sites are described. The changes in the study design were primarily due to the nonexistence of an ideal measure of English language proficiency (MELP), which RTI had assumed did exist and would be used during the field test. Since the ideal MELP did not exist, the study design was changed from an area household interviewing approach to the interviewing of elementary school children and adults who had been identified by school districts as either limited English-speaking ability (LESA) or non-LESA persons. The categorization of sample individuals introduced an external criterion, namely school district determined LESA/non-LESA status, into the analysis. It was felt by some that this external criterion was, if not an ideal MELP, at least one of several highly promising MELPs. The other potentially ideal MELPs resulted from (1) scores on paper and pencil tests developed by CAL and (2) a five point scale rating of the respondent's English language proficiency performed by the field interviewer.

B. Original Proposed Design

One objective of the field test was to test one or more surrogate methods of determining or measuring an individual's proficiency in the English language. The surrogate measure of English language proficiency (MELP) were to be validated using an ideal measure of MELP. The ideal MELP, however, would involve the use of paper and pencil tests and other features unacceptable to the Bureau of the Census, as far as utilizing the ideal method in a survey conducted by the Bureau. RTI visualized that the purpose of the field test was to identify a surrogate method of determining MELP which was

sufficiently accurate and which was sufficiently simple in form that it could be included in the Census survey.

The contemplated approach for validating the abbreviated MELP involved administering the abbreviated MELP, followed by the ideal MELP, to samples of several different bilingual ethnic groups. Having obtained these data, the discriminatory power of the abbreviated MELP alternatives could then be analyzed and compared with the results obtained by the ideal MELP.

More specifically, RTI's Sampling Department was to select several sites containing concentrations of ethnic groups thought to have limited English-speaking ability. RTI staff members would then travel to each site and identify specific areas with high concentrations of the populations of interest. Data would be obtained for approximately 250 young people aged 5 to 17 years and for approximately 250 adults aged 18 and older within each selected site. It was anticipated that approximately 450 ethnic group households per site would have to be surveyed to obtain the above numbers of completed cases. Considerably larger numbers of households would be screened in order to identify 450 ethnic group households; the total number screened was expected to vary from one site to another. In selecting field test sites and areas within sites, attempts would be made to balance the selection of rural and urban areas, although it was expected that most of the fieldwork would be done within low-socioeconomic areas. An incentive payment of $5.00 per respondent was recommended. For those individuals under the age of 10, it was expected that two 30-minute data collection sessions would be used rather than one longer session.

C. Revised Design

The work performed by RTI under subcontract to CAL was begun in the first week of June based on acceptance of RTI's proposal number 250-75-06-02 of May 26, 1975 as modified by RTI's June 2, 1975 letter proposal number 250-75-06-03. The technical approach and field test sites had been discussed extensively by RTI and CAL, and in turn by CAL and NCES, so that further changes in the project work plan or the sites were not expected after June 2. In fact, the sites had already been revised at least two times in response to CAL/NCES questions on RTI's first and second proposals. There had never been any questions raised with RTI by CAL/NCES concerning the general approach outlined in the proposal--that is, defining clusters of housing units where concentrations of the ethnic-linguistic group of interest resided (within each site) and conducting personal interviews with selected members of the sample households. RTI staff, upon receiving notice from CAL that our proposal was accepted, began newspaper advertising and sent RTI staff to begin recruiting field interviewers in the seven field test sites the week of June 16-20. During the same week, an RTI staff member traveled to San Francisco to observe the CAL formative work in developing the MELP instruments. Upon arriving in San Francisco, it was found that CAL/NCES were in the process of revising the project work plan and that none of the seven field test sites could then be considered firm.

As a result, the approach was changed from the household interviewing approach proposed to interviewing of elementary school children and adults defined as limited English-speaking ability (LESA) and non-LESA persons by screening programs within the sites. It was agreed at the San Francisco meeting June 20 that CAL/NCES would be responsible for identifying field

test sites and for acquiring lists of sample members to be interviewed. At that meeting RTI pointed out that any delay in identifying new sites would have both schedule and cost implications.

D. Selection of Sites

RTI's original recruiting schedule called for project personnel to visit during June 23-27 the seven original sites: Miami, New York City, San Antonio, Manchester, N.H., Lafayette, La., San Francisco, and Apache County, Arizona. While recruiting in these sites was in progress, RTI received word from NCES at various times during the period from June 23 to July 9 as follows: (1) four of the original seven sites - New York City, Manchester, Lafayette, and San Antonio - were no longer being considered, (2) Apache County was in doubt, (3) Miami, San Francisco, and El Paso were "probables," (4) Camden, N.J. was firm, (5) Apache County, Miami, and El Paso were firm, (6) Camden and San Francisco have refused, (7) Camden agreed to participate, and (8) San Francisco agreed to participate. Accordingly, RTI took the following actions: (1) recruiting in the four dropped sites was cut short (although it was too late to effect any appreciable cost savings), (2) recruiting in Apache County was allowed to continue pending more definitive word as to its fate, (3) arrangements were made to have a staff member recruit in the two new sites, Camden and El Paso. On July 29, Camden refused for the final time.

The lists were received during the week of July 14-18. The change in the field test design, and the resulting changes in the sites, delayed the field-work one week in two field test sites. The change in the work plan disrupted and delayed other planned activities, such as work on the

instrument and preparation of training materials. However, the field test was finally carried out approximately two weeks behind schedule according to the revised plan.

III. SAMPLING ACTIVITIES

A. Securing School District Cooperation

RTI's original proposal specified seven sites in which field test data were to be collected using a household survey approach. Once the field test design had been modified to a list sampling approach, the question of which sites to use in the field test was thrown open again. That is, in addition to considerations of ethnic-linguistic groups and geographic representation, it was necessary to identify school districts within the sites which had exemplary screening programs for bilingual education programs and which would also cooperate with RTI/CAL/NCES by providing lists of children who had been screened and found to be LESA or non-LESA. Lists of adults who were or had been enrolled in English for Speakers of Other Languages were also searched out, but the primary emphasis was on securing cooperation and lists of classified elementary school children.

NCES made numerous telephone contacts in order to locate school districts which had the desired characteristics and which would cooperate. This proved to be a much more time consuming process than anticipated and resulted in using three of the originally planned seven sites--Miami, San Francisco, and Northeast Arizona--in the field test. Although many additional sites were considered and contacted during this period, it was finally decided that El Paso, Texas and Camden, N.J. also would be included in the field test. Unfortunately, cooperation was never obtained from Camden although a definite "no" response was not obtained from the Camden school district until July 29, at which time it was considered too late to add a new site to the field test. Thus, the field test was carried out in four sites--Miami, El Paso, Northeast Arizona, and San Francisco.

B. Obtaining Lists of Students

After the school districts had been identified and had agreed to participate, RTI discussed with school district personnel the requirements for constructing lists of students or duplicating existing lists if such existed. In general, the information needed included the student's name, address, grade, and LESA classification. In some cases, parent's name, address, ethnic group, and telephone number were also easily available and were obtained. School districts in each site were asked to provide names of at least 500 children who had been screened and classified, of which approximately one-half had been determined to be LESA. In addition to providing specifications concerning the lists to be constructed, RTI also provided assistance to the school districts in preparing the lists when this was needed.

In Miami, an RTI staff member spent several days constructing the lists of children and adults in cooperation with school district personnel. In El Paso, RTI paid the salary of a person hired for one week by the school district to construct the desired lists of children and adults. An RTI staff member also worked with the Ganado and Window Rock, Arizona school districts for approximately one week constructing the lists of children (no lists of adults were obtained in Arizona or San Francisco). The San Francisco school district prepared their own lists of children and gave the lists directly to NCES.

C. Selection of Sample Individuals

In general, samples of children and adults were selected from two or more lists in each site (see Table III.1). Several controls were exercised

in order to meet the project objectives, as follows:

- Controls to insure that approximately equal numbers of interviews were completed with children and adults,
- Controls to insure that approximately equal numbers of interviews were completed from each of the lists sampled,
- Controls to increase the precision of comparisons between bilingual and monolingual interviewers, and
- Controls to randomize the subsample of interviews which were monitored by CAL staff.

Lists had been obtained of elementary school children who had been screened by their school districts and classified as to their English-speaking ability. The lists included each student's name, address, grade, and LESA classification. In some cases, parent's name, birthdate, ethnic group, and telephone number were also available. In Miami and El Paso, lists of adults who were enrolled or had been enrolled in classes of English for speakers of Other Languages had also been obtained.

The lists obtained were stratified by list code, age group, and site. Initial samples of approximately 500 names per age group per site were selected randomly within strata such that approximately equal samples were obtained from each list code stratum within sites. (In San Francisco and Northeast Arizona, adults were randomly sampled from the households in which children had been selected.) After beginning the field work, supplemental samples were selected in some sites in order to increase the number of interviews completed over the previously established target number of 250 per age group per site. Table III.1 shows the number of names on each of the lists obtained from each school district and the total number of sample

Table III.1 Sample Sizes by Site, Age Group, and List Code Strata

Site	Age Group	List Code (LESA Classification)	Total number on list	Total number selected
Miami	Children	Non-independent	320	213
		Intermediate	531	248
		Independent	645	230
		Total	1,496	691
El Paso	Children	Spanish-dominant	305	305
		English-dominant	293	293
		Total	598	598
San Francisco	Children	Limited English-speaking	396	343
		Non-Limited	314	297
		Total	710	640
N.E. Arizona:				
Ganado	Children	Non-Independent	144	144
		Intermediate	434	142
		Independent	237	145
		Total	815	431
Window Rock	Children	Reading below grade level	623	143
		Reading at or above grade level	518	145
		Total	1,141	288
Total (N.E. Arizona)	Children	Total	1,956	719
Miami	Adults[1/]	Beginner	246	246
		Intermediate	172	172
		Advanced	282	282
		Total	700	700
El Paso	Adults	Beginner	222	222
		Advanced	278	278
		Total	500	500

1/ Note no lists of adults were obtained in Arizona and San Francisco.

names made available to the field staff, by stratum within site. It should be noted that, in some cases, not all of the names selected were assigned to interviewers. An initial assignment was given to each interviewer and the remaining names were assigned as needed to replace those who refused or could not be located. Each interviewer was assigned some cases from each of the list code strata sampled within the site. A random subsample of approximately one-fourth to one-third of each interviewer's assignment was designated for monitoring by CAL staff members. In addition, five pairs of interviewers were assigned to work in five separate areas within each site. Each pair consisted of one monolingual and one bilingual interviewer; these interviewers were randomly chosen to participate in this substudy. The assignment of sample cases to each pair member was randomized within each area. This was done to increase the precision of comparisons between monolingual and bilingual interviews.

IV. DATA COLLECTION ACTIVITIES

A. Overview

Data collection activities for the Bilingual Survey involved interviews with over 2,700 children and adults from non-English-language backgrounds in four sites: Miami (Cubans), El Paso (Chicanos), N. E. Arizona (Native Americans), and San Francisco (Asians, predominantly Chinese). Respondents were selected from lists provided by local school districts. In all four sites lists of children from the ethnic groups of interest were provided; about half of the children listed for each site had been administratively classified by the local school district as having limited English-speaking ability (LESA), while the other half had been classified as not being limited by reason of proficiency in English (non-LESA). In addition, the school districts in two sites (Miami and El Paso) provided similar lists for adults enrolled in adult basic education programs; in the other two sites (N. E. Arizona and San Francisco) adult respondents were selected at random from the households of child respondents. In each site two RTI staff members supervised a team of approximately 25 interviewers during the fieldwork period.

During the period from June through early September 1975 data collection activities were associated with the following principal tasks:

- design of the data collection plan;
- consultations with CAL on instrument development;
- recruitment of interviewers;
- final preparations for training and data collection;

- training of project personnel;
- conduct of the fieldwork;
- data receipt, edit, and reduction; and
- attendance at Language Group Representative (LGR) meetings.

The execution of these tasks and the results obtained are discussed in subsequent sections of this chapter. A chart showing the period of performance for principal data collection activities is presented in Table IV. 1.

B. Design of the Data Collection Plan

As described in detail in Chapter II, RTI's original data collection plan underwent substantial revision during June 1975 as a result of a change in the study design. To summarize, during the instrument development phase of the project, CAL/NCES decided that the field test data should be compared with determinations by school districts as to whether an individual was LESA or Non-LESA. (The school districts use the LESA classification to determine if an individual should be enrolled in special programs.) Accordingly, the data collection plan was changed from RTI's proposed approach involving household interviews in areas with a concentration of one of the ethnic groups of interest, to one involving interviews with elementary school children and adults defined as LESA and Non-LESA by screening programs conducted by local school districts within the field test sites.

It was understood that CAL/NCES would be responsible for revising the list of field test sites, as necessary, and for obtaining the cooperation of local school districts in providing appropriate lists of LESA and Non-LESA children and adults. The list of sites underwent a number of revisions in late June and early July, but by mid-July,

Table IV.1

Schedule of Principal Data Collection Activities

		Week Beginning Monday: 6/2	6/9	6/16	6/23	6/30	7/7	7/14	7/21	7/28	8/4	8/11	8/18	8/25	9/1
Attend Spanish and Native American LGR meetings at CAL	(D. DeWitt)		—												
Consult with CAL staff in San Francisco on instrument development	(N. Weeks)			—			—								
Attend European LGR meeting at CAL	(P. Moore)			—											
Attend second LGR meeting in San Francisco	(D. Horvitz and N. Weeks)							—							
Recruit interviewers in Miami	(A. Davis)				—	—									
Recruit interviewers in New York City	(L. Weiss)				—										
Recruit interviewers in San Antonio	(D. Grossman)				—										
Recruit interviewers in Manchester	(D. Boytos)				—										
Recruit interviewers in Lafayette	(J. Shirey)				—	—									
Recruit interviewers in San Francisco	(T. McAdams)				—	—									
Recruit interviewers in N.E. Arizona	(J. Durham)				—	—	—	—							
Recruit interviewers in Camden	(D. Boytos)					—	—	—	—						
Recruit interviewers in El Paso	(D. Grossman)					—									
Visit Miami to obtain lists from school district	(N. Thorne)						—								
Complete preparations for training and data collection (finalize and reproduce manual, forms, and instruments, obtain necessary supplies; complete sampling from lists; etc.)								—							
Central staff briefing session								—					—		
Site team/interviewer training (Miami and El Paso)									—						
Site team/interviewer training (San Francisco and N.E. Arizona)										—					
Data collection (Miami and El Paso)									—	—	—	—			
Data collection (San Francisco and Arizona)										—	—	—	—		
Data receipt, edit, and reduction											—	—	—	—	—
Attend final LGR meeting in Roslyn, Va.	(selected RTI staff)														—

school lists were obtained from four sites ultimately included in the field test. As previously noted, lists of LESA and non-LESA children were obtained from each site; however, corresponding lists for adults were not available for N. E. Arizona and San Francisco, which necessitated the random selection of adult respondents from the households of child respondents in these sites.

The change in the data collection plan and subsequent revisions in the list of field test sites resulted in RTI's incurring some unanticipated costs, primarily associated with recruiting interviewers in new sites. In addition, recruiting costs were incurred in five sites that were later dropped. As noted in Chapter II it was also necessary to postpone the commencement of data collection one week in N. E. Arizona and San Francisco.

C. Instrument Development

CAL was responsible for developing the instruments used in the field test and for preparing appropriate instructions for instrument administration. RTI consulted with CAL during the instrument development phase of the project and was responsible for formatting the Census Questionnaire and for printing this and other instruments developed by CAL.

The instrument development work was conducted by CAL, in consultation with NCES and RTI, in San Francisco from early June through mid-July. During this period an RTI staff member made two trips to San Francisco in order to offer input from a data collection viewpoint. The instruments developed and subsequently administered in the field test are listed and briefly described below:

1. The Census Questionnaire (CQ) - The CQ contained 32 candidate Census-type items and was administered to all designated

respondents (except children aged 9 or younger) as well as to household respondents answering on behalf of designated respondents. An accompanying set of flash cards was also used in the administration of this instrument.

2. The Household Information Form (HIF) - The HIF contained 9 items similar to items on the Census Current Population Survey instrument (July 1975 version) and obtained data on the household (number of members, facilities, etc.) and the education and employment of the head of household. The HIF was administered to a household respondent in each sample household. The set of flash cards used for the CQ were also used in the administration of this instrument.
3. The Mat-Sea-Cal (MSC) - The MSC was administered to designated child respondents (DCRs) in order to measure a child's ability to comprehend spoken English and to use English orally to answer questions. The DCR was asked to point out pictures that corresponded to orally administered statements, and to answer specific questions about other pictures.
4. The Adult Production Test (APT) - The APT was administered only to designated adult respondents (DARs) in order to evaluate the ability of the DAR to use English orally. The DAR was asked to make statements, and ask and answer questions about a set of pictures.
5. The Adult Comprehension Test (ACT) - The ACT was also administered only to DARs and was used to measure the DAR's ability to comprehend spoken English. The DAR was asked to

listen to pairs of sentences and to judge whether the two sentences were the same or different in meaning.

6. The Oral Communications Test (OCT) - The OCT was administered to both DCRs and DARs and measured a respondent's ability to communicate in English. The respondent was asked to provide a simple description of a picture which was randomly selected from a set. The interviewer then made a "guess" as to which picture the respondent had described.

7. Direct Rating Scales - Initially, a procedure was developed and implemented that provided a direct measure of a respondent's understanding of and speaking proficiency in English. Administered to all respondents at the conclusion of the interview, the procedure called for the interviewer to ask the respondent three conversational questions designed to elicit a free response, and then to rate his understanding and speaking proficiency on a five-point scale. (This scale is referred to as the DORT in Chapter V.) Shortly after the fieldwork began, two more rating scales were added: a rating of the interviewer's ability to understand the respondent's English, and a rating of the respondent's apparent ability to understand the interviewer's English. The latter two ratings also involved a five-point scale but there were no questions to ask; rather, the interviewer was free to consider all of the interaction with the respondent in determining the appropriate rating on each scale.

For the analysis described in Chapter V the results of the MSC and the OCT for children were combined to obtain a MELP score called the CP (criterion package) score. In particular, the CP score for children was obtained by combining the results from 32 items on the MSC and 15 items on the OCT. Similarily, for adults a CP score was obtained by combining the results from 16 items on the APT, 15 items on the OCT and 10 items on the ACT.

In addition to developing the instruments described above, CAL staff also provided RTI with written instructions for administration of the criterion measures.[1/] As previously noted, RTI worked closely with CAL during the instrument development period and offered aid and advice on data collection aspects of the instruments. RTI was also responsible for designing and reproducing the instruments and related materials (picture booklets, answer sheets, test keys, etc.) and for writing instructions covering the administration of the CQ and HIF. These instructions, together with CAL's instructions for the administration of the criterion measures, were included in a field manual developed by RTI covering all aspects of data collection. A complete set of all instruments and manuals is included as an attachment to this report.

D. Recruitment of Interviewers

The RTI site supervisor for each field test site was responsible for

---

1/ The criterion measures include the MSC, APT, the ACT, the OCT, and Direct Rating Scale.

recruiting an appropriate interviewing staff to conduct data collection. Every effort was made to develop leads on qualified bilingual prospects. Prospective applicants were obtained primarily from classified newspaper ads, although other sources of qualified applicants used included RTI's National Interviewer File and contacts with other survey research organizations and the U. S. Bureau of the Census. The site supervisors visited each site during late June and July to interview applicants personally, check work references on the top applicants, and retain the most qualified. As indicated in Table IV.1, recruiting took place in nine sites, five of which were subsequently dropped due to revisions in the data collection plan.

Table IV.2 summarizes data on RTI's recruiting effort in the four sites where the field test was carried out. As shown in that table, the site supervisors received a total of 299 telephone inquiries from the newspaper ads. These prospects were screened on the telephone and personal interviews were scheduled with the best qualified as well as with leads developed from other sources. Of the total of 285 personal interviews conducted by the site supervisors in the four field test sites, 101 interviewers were retained. Of these 51 were bilingual (i.e., they spoke English and also the language of the target population) and 50 were monolingual (i.e., they spoke English but did not speak the language of the respondent). As evident from the table, bilingual interviewers with experience in survey interviewing were difficult to find. Of those interviewers with no related experience, preference was given to those who were bilingual.

Table IV.2

DATA ON RECRUITING INTERVIEWERS FOR THE FIELD TEST

	Miami	El Paso	Arizona	San Francisco	Total
Telephone Inquiries from News Ads	123	65	3*	108	299
Applicants Interviewed	76	51	71	87	285
Bilingual	45	32	33	15	125
Experienced in Survey Interviewing	6	3	2	1	12
No Experience in Survey Interviewing	39	29	31	14	113
Monolingual	31	19	38	72	160
Experienced in Survey Interviewing	7	12	5	22	46
No Experience in Survey Interviewing	24	7	33	50	114
Interviewers Retained	25	23	23	30	101
Bilingual	15	14	13	9	51
Experienced in Survey Interviewing	6	3	1	1	11
No Experience in Survey Interviewing	9	11	12	8	40
Monolingual	10	9	10	21	50
Experienced in Survey Interviewing	3	4	2	13	22
No Experience in Survey Interviewing	7	5	8	8	28

Percentage of available experienced interviewers retained 57%

Percentage of available inexperienced interviewers retained 30%

Percentage of available bilingual interviewers retained 41%

Percentage of available monolingual interviewers retained 38%

* Four ads were run, only one of which instructed applicants to inquire by telephone. The other three ads instructed applicants to apply in person.

NOTE: All interviewers spoke English. For purposes of this study, "monolingual" referred to interviewers who did not also speak the language of the respondent, while "bilingual" interviewers did speak the respondent's language.

E. Final Preparations for Training and Data Collection

During the week of July 14 RTI project staff completed preparations for field training and data collection. These activities included:

- developing a field manual for interviewers covering all aspects of the fieldwork;
- preparing written instructions for the site supervisory team;
- finalizing the Census Questionnaire and Household Information Form;
- finalizing the materials associated with the criterion measures (picture booklets, answer sheets, test key, Mat-Sea-Cal sketch, etc.);
- designing appropriate project field forms (Record of Household Contacts, Receipt for Payment, Case Control Form, etc.);
- reproducing appropriate quantities of field materials (manual, questionnaires, materials associated with the criterion measures, field forms, etc.);
- obtaining necessary field supplies (e.g., clipboards, portfolios, pencils, RTI return envelopes, ID cards, RTI Interviewer's General Manual, etc.); and
- sampling from the school lists and preparing for each sample person a 3" x 5" card containing the name, address, ID number, and other data.

On July 18 a briefing session was held at RTI for central staff who were to participate in field training beginning the following week. This

session was also attended by two CAL project personnel. The session was conducted by the RTI staff member responsible for the data collection effort and included a review of all training procedures, materials, techniques, and aids, as well as a discussion of the kinds of questions likely to be raised by the trainees.

F. Training of the Field Staff

Field staff training was conducted during the week of July 21 in Miami and El Paso and during the week of July 28 in N. E. Arizona and San Francisco. The initial schedule called for training to take place in all field test sites during the week of July 13; however, early in the contract period RTI decided to push the proposed training and data collection schedule back one week in order to allow more time to perform necessary preparatory tasks. As pointed out previously, it later became necessary to postpone training and the commencement of data collection in N.E. Arizona and San Francisco an additional week due to delays in obtaining school lists from those sites.

A three-day interviewer's training session was conducted in each of the four field test sites, convening at 9:00 a.m. on Tuesday and adjourning on Thursday afternoon. Each session was under the direction of an RTI central staff senior survey specialist, assisted by the site supervisor and assistant site supervisor for each site. A team of three or four CAL project staff members was also present at each session and provided assistance to the RTI trainers, especially with the presentation of the criterion measures.

During each session the field manual prepared specifically for the field test and RTI's Interviewer's General Manual, which covers general

topics and standardized procedures applicable to all RTI surveys, were covered in detail. Special emphasis was given to training the monolingual interviewers in establishing rapport with the target ethnic group and to reviewing and discussing the problems likely to be encountered in the administration of the field test instruments to persons with limited proficiency in English. The training sessions included instructor demonstrations, group discussion and interaction, supervised classroom practice, and finally, a quiz on all field procedures. Satisfactory performance on the quiz and in classroom practice was mandatory; interviewers who did not perform at an acceptable level were given additional training or released, at the discretion of the trainers.

Data collection in each field test site was supervised by two RTI staff members, a site supervisor and an assistant site supervisor, who remained "on-site" during the data collection period. The two supervisory personnel in each site were briefed by the senior survey specialist in charge of the interviewer training session on Monday prior to the commencement of the session on Tuesday. All supervisory and interviewer procedures were covered in detail during this briefing session and plans for the conduct of interviewer training were finalized.

G. Data Collection Procedures

Data collection was begun in each site immediately after the interviewer training session and continued for approximately three weeks. The two RTI supervisors in each site directed the data collection activities of from 23 to 30 interviewers. In addition, three or four CAL staff members were in each site during the data collection period and monitored approximately 15 percent of the field interviews.

In order to compare the effect of bilingual versus monolingual interviewers, a substudy was conducted that involved matching five bilingual interviewers with the same number of monolingual interviewers in each site and making random work assignments to each group. The results of this substudy are discussed in Chapter V.

Interviewer assignments were prepared by the site supervisory teams, following detailed procedures designed by RTI's Sampling Department to (1) equalize the effort for children and adults; (2) equalize the effort for each child or adult's proficiency level defined by the schools (e.g., in Miami: non-independent, intermediate, and independent); (3) increase the precision of the comparison between bilingual and monolingual interviewers; and (4) randomize the subsample of interviews to be monitored by the CAL staff.

The field procedures followed by the interviewers during the field test are detailed in the interviewer's field manual, a copy of which is included in the attachment to this report. The procedures for the three principal types of cases are summarized below:

. Designated Child Respondents (DCRs)

(1) The interviewer calls in person at the sample household at a time when a household respondent (household member at least 14 years old) is likely to be home.

(2) The interviewer locates a household respondent and (a) introduces herself, (b) verifies that the DCR is a household member, and (c) explains the study.

(3) The interviewer administers the Census Questionnaire (CQ) and Household Information Form (HIF) to the

Household Respondent. (NOTE: The household respondent responds to the CQ on behalf of the DCR.)

(4) The interviewer determines the age of the DCR.

(5) The interviewer interviews the DCR. (NOTE: If the DCR is ten or older, the interviewer administers the CQ and criterion measures; if the DCR is nine or younger, the interviewer administers only the criterion measures.)

. Designated Adult Respondents (DARs) from School Lists (Miami and El Paso)

(1) The interviewer locates a household respondent as for DCRs above.

(NOTE: The household respondent can also be the DAR, if the DAR is the first person 14 or older the interviewer encounters.)

(2) The interviewer administers the CQ and HIF to the household respondent.

(NOTE: The CQ is second-hand if the household respondent is not also the DAR; first-hand if the household respondent is the DAR.)

(3) The interviewer interviews the DAR.

(NOTE: If the household respondent is the DAR, the CQ will have already been administered and the interviewer continues with the criterion measures.)

. Designated Adult Respondents (DARs) Randomly Selected from DCR Households (N.E. Arizona and San Francisco

The interviewer locates a household respondent, as above. The interviewer then randomly selects an adult member of

the household, who becomes the DAR. The interviewer then proceeds to interview the household respondent, DCR, and DAR as described above.

A number of minor procedural changes and refinements were made as the fieldwork progressed and problems became apparent. One noteable change that was implemented near the end of the fieldwork period concerned obtaining second-hand CQ information on adults. In order to increase the number of cases where second-hand CQ data were obtained on DARs, interviewers were instructed to attempt to find a household respondent who was not also a DAR. One callback was authorized to accomplish this, if necessary.

Respondents were paid cash incentives by the interviewers at the rate of $2.00 for each completed CQ and $2.00 for each completed set of criterion measures. Incentive payments made directly to DCRs were made with the knowledge of a responsible adult member of the household. No payment was made for the short HIF, which was completed in conjunction with the initial CQ.

Interviewers were instructed to make up to two calls at a sample household in order to contact a household respondent. If the interviewer was unable to contact a household respondent on the first call, she would attempt to find out from neighbors when the household residents were most likely to be found at home, and made her second call at that time. If neighbor information was unavailable, the interviewers were instructed to make the return call after 6:00 p.m. on a weekday or on a weekend. After initial contact, the interviewer was allowed up to two more calls to complete interviewing in the household. If she had still not completed

her work at the household after two additional callbacks, she was instructed to discuss the case with a site supervisor immediately.

The interviewers were not permitted to substitute non-sample persons for designated respondents. All non-interview cases had to be discussed with a site supervisor, who would determine what, if any, additional action should be taken. If no further action was warranted, the supervisor would approve the noninterview result and provide the interviewer with a substitute case, according to the interviewer assignment procedures developed by RTI's Sampling Department.

The two RTI supervisors in each site remained in the field during the fieldwork period in order to monitor closely the data collection activities of the interviewers. The supervisors normally met with each interviewer at least twice a week to review the status of each of her active cases and to advise and assist her as necessary. The supervisors were responsible for editing and approving the instruments associated with each completed case and for mailing completed cases to RTI on a flow basis. Additional cases were assigned to interviewers when appropriate, following procedures specified by RTI's Sampling Department. The supervisors were also responsible for validating the fieldwork by contacting at least ten percent of each interviewer's respondents (those not monitored by CAL staff) to verify that the interviewer had conducted the interview properly and that the respondents had been paid. Other responsibilities of the site supervisors included monitoring interviewer costs; controlling the issuing and retrieving of advances to interviewers for use in making cash payments to respondents; recruiting and training replacement interviewers, as necessary; maintaining records on the handling and status of each case; and reporting to RTI at least weekly the status of the fieldwork in the field test site.

H. Fieldwork Results

Data collection was completed on August 16 in Miami and El Paso and on August 23 in N.E. Arizona and San Francisco. The results of the fieldwork are summarized in Table IV.3. This table shows results for each of the four field test sites as well as the composite results for all sites. These data were obtained manually from interviewer and supervisor records. The figures were not verified using computer-generated data since most of the data are not available in machine-readable form.

As the table indicates, a total of 2,704 respondents were administered the appropriate criterion measures: 1,472 of these were children and 1,232 were adults. The respondents interviewed represent 63 percent of the potential respondents assigned; of the 1,615 nonrespondents (37 percent), 114 (3 percent) refused to be interviewed and the remaining 1,501 nonrespondents (35 percent) involved cases where the sample members were unavailable for a variety of reasons (see footnote 3 to the table). The time and mileage figures shown were well within budget constraints, and the percent of DARs for whom second-hand CQ data were obtained was slightly lower than expected.

Table IV. 4 contains some comparison data on data collection results for monolingual and bilingual interviewers. These figures seem to indicate that the monolingual interviewers were not at a disadvantage vis-a-vis bilingual interviewers; in fact, the former performed slightly better, overall, than did the latter, probably because more of the monolingual interviewers had previous experience in survey interviewing. Additional comparison data will be presented Chapter V.

Table IV.3

DATA COLLECTION RESULTS OF FIELD TEST[1/]

	Miami	El Paso	Arizona	San Francisco	Total
Potential Respondents Assigned[2/]	1,079	1,071	972	1,192	4,314
Interviews with Children	335	426	356	353	1,472
Interviews with Adults	333	265	315	319	1,232
Total Interviews (Percent)	668 (62%)	691 (65%)	673 (69%)	672 (56%)	2,704 (63%)
Refused (Percent)	26 (2%)	18 (2%)	16 (2%)	54 (5%)	114 (3%)
Other Nonrespondents[3/] (Percent)	385 (36%)	362 (34%)	283 (29%)	471 (40%)	1,501 (35%)
Total Nonrespondents - (Percent)	411 (38%)	380 (35%)	299 (31%)	525 (44%)	1,615 (37%)
Total Hours Charged[4/]	2,916	2,992	3,203	2,917	12,028
Total Miles Driven[5/]	22,966	21,079	34,328	8,299	86,672
Average Hours Per Interview	4.4	4.3	4.8	4.3	4.5
Average Miles Per Interview	34.4	30.5	51.0	12.4	32.1
% of Adult Respondents with 2nd Hand Census Questionnaires[6/]	36%	36%	83%	36%	48%

1/ Figures in this table are based upon manual counts and computations by interviewers and supervisors and have not been verified by machine tabulations.

2/ In Miami and El Paso both children and adults were assigned to interviewers. In Arizona and San Francisco only children were assigned, since no adult lists were obtained for these sites. Interviewers randomly selected an adult from each sample child's household in these sites. For Arizona and San Francisco, therefore, the number of potential respondents was twice the number of sample children assigned.

3/ Examples of "other" nonrespondents include cases where the sample member had moved to another city; where the address was nonexistent; where the sample member could not be contacted at home in the prescribed number of interviewer visits; where the sample member was out of town; or where he was sick, institutionalized, or otherwise unavailable.

4/ Includes training time.

5/ Includes mileage incurred in connection with training.

6/ Figures shown indicate the percent of adult respondents in each site about whom Census Questionnaire data were obtained from a household member other than the respondent as well as from the respondent himself.

Table IV.4

COMPARISON OF DATA COLLECTION RESULTS FOR MONOLINGUAL AND BILINGUAL INTERVIEWERS[1/]

	Miami		El Paso		Arizona		San Francisco		Total	
	Mono.	Bi.	Mono.	Bi.	Mono.	Bi.	Mono.	Bi.	Mono.	Bi.
No. of Interviewers[2/]	10	15	9	14	10	13	21	9	50	51
Potential Respondents Assigned[3/]	404	675	397	674	533	439	803	389	2137	2177
Respondents Interviewed (Percent)	249 (62%)	419 (62%)	260 (65%)	431 (64%)	394 (74%)	279 (64%)	470 (59%)	202 (52%)	1373 (64%)	1331 (61%)
Refused (Percent)	11 (3%)	15 (2%)	7 (2%)	11 (2%)	8 (2%)	8 (2%)	39 (5%)	15 (4%)	65 (3%)	49 (2%)
Other Nonrespondents[4/] (Percent)	144 (36%)	241 (36%)	130 (33%)	232 (34%)	131 (25%)	152 (35%)	294 (37%)	172 (44%)	699 (33%)	797 (37%)
Total Nonrespondents (Percent)	155 (38%)	256 (38%)	137 (35%)	243 (36%)	139 (26%)	160 (36%)	333 (41%)	187 (48%)	764 (36%)	846 (39%)
Total Hours Charged[5/]	1099	1817	1162	1810	1717	1486	1937	980	5935	6093
Total Miles Driven[6/]	9412	13554	8106	12973	18929	15369	7088	1011	43535	43137
Average Hours Per Interview	4.4	4.3	4.6	4.2	4.4	5.3	4.1	4.9	4.3	4.6
Average Miles Per Interview	37.8	32.4	31.2	30.1	48.0	55.2	15.1	6.0	31.7	32.4

1/ Figures in this table are based upon manual counts and computations by interviewers and supervisors and have not been verified by machine tabulations.

2/ All interviewers spoke English. For purposes of this study, "monolingual" referred to interviewers who did not also speak the language of the respondent, while "bilingual" interviewers did speak the respondent's language.

3/ In Miami and El Paso both children and adults were assigned to interviewers. In Arizona and San Francisco only children were assigned, since no adult lists were obtained for these sites. Interviewers randomly selected an adult from each sample child's household in these sites. For Arizona and San Francisco, therefore, the number of potential respondents was twice the number of sample children assigned.

4/ Examples of "other" nonrespondents include cases where the sample member had moved to another city; where the address was nonexistent; where the sample member could not be contacted at home in the prescribed number of interviewer visits; where the sample member was out of town; or where he was sick, institutionalized, or otherwise unavailable.

5/ Includes training time.

6/ Includes mileage incurred in connection with training.

I. Processing Completed Field Test Instruments

After editing and approving completed field test instruments, the site supervisors mailed them, batched by households, to RTI on a flow basis. Upon receipt at RTI, they were routed to a receipt control station, where the survey instruments for each household were logged in manually by ID number.

From the receipt control station the documents were routed to the scan-edit/coding station, where each document was scan-edited on an item-by-item basis and appropriate codes were assigned, where required. A CAL staff member who visited RTI for several weeks during late July and August was instrumental in the development of detailed scan-edit procedures and questionnaire codes.

After the instruments passed through the check-in and scan-edit/coding stations, they were encoded at the direct data entry station. Direct data entry means simply that programmable terminals with keyboards and cathode ray tube display screens are used instead of manual coding and keypunch. A programmed format specifically designed for the survey questionnaire appears on the display screen. As the terminal operator looks at the hard-copy document, the responses are keyed into the terminal. The keyed data then appear instantaneously on the display screen and are recorded on magnetic tape cassettes for transmission to a computer facility.

After transmission to the computer terminal, a quality control check was made by twice rekeying a six percent random sample of the transmitted instruments. The keystroke error rate was then calculated and showed an error rate of less than 0.6 percent.

Following data reduction, all hard copy documents that could be used

to identify respondents were shredded. Other documents are being stored on the RTI campus.

J. Confidentiality Procedures

As with all Institute projects, RTI exercised extreme care during data collection and processing to insure that no one had access to the identity of respondents except authorized RTI and CAL project staff.

Confidentiality requirements were emphasized to the supervisors and interviewers during training. They were instructed that completed or unused questionnaires were not to be given to anyone not involved in the survey; that duplication of materials was not permitted; and that the identity of respondents and their answers must be kept confidential at all times.

At RTI access to completed survey instruments was carefully controlled during all stages of processing. Overnight storage of processing batches was provided in a locked and secure work area. When processing was completed, hard copy documents were filed in an ordered, accessible manner and maintained in a secure, well-ventilated vault-like room in the basement of the Ragland building on the RTI campus. Access to this room is by one locked door, with controlled (i.e. signed for) entry only on authority of the survey specialist in charge of data processing operations. As previously noted, in September all hard copy documents that could be used to identify respondents were pulled from storage and shredded.

K. Attendance at Language Group Representative (LGR) Meetings

At CAL's invitation, RTI staff attended and participated in a number of the LGR meetings scheduled by CAL. In June each of the five LGR groups met individually at CAL and an RTI staff member was present for at

least part of three of those sessions (Spanish, Native American, and European.) All the LGR's met as a group in San Francisco July 13-14 and again in Roslyn, Virginia on September 3-4. Selected RTI project staff members attended both of these sessions. RTI project staff found these sessions to be fruitful and a number of the suggestions made by the LGRs were incorporated into RTI's data collection and analysis plans.

V. ANALYSIS ACTIVITIES

A. Overview

This chapter describes the analysis performed by RTI on the data collected and processed by RTI as described in Chapter IV. This analysis was carried out only after detailed consultation between RTI and CAL staff, and all RTI analysis activities were directed towards helping CAL make recommendations to NCES on (i) the best method of using Census Questions to measure English language proficiency and (ii) the procedures to be used by the Census in their Survey of Income and Education in collecting MELP data. Analysis activities by RTI began in September 1975 and are still continuing. Throughout the analysis phase of the project, RTI and CAL staff were in almost daily contact and RTI performed a great many specific analyses and computer runs at the direct request of CAL. In many cases, RTI did not analyze specific computer runs but mailed the runs directly to CAL for their analysis. This chapter only presents analyses performed by RTI and does not attempt to discuss CAL analyses based upon computer runs performed by RTI.

As mentioned previously, the analysis of the data was made much more difficult due to the nonexistence of an "ideal" MELP to use as a standard for comparison with the potential MELPs based upon the Census Questions. This resulted in a number of analyses to help determine which of the several "ideal" MELP's; e.g., CP (criterion package) score, LIST ... was best in some sense and where to divide these measures to determine LESA/Non-LESA groups.

In this chapter, Section B gives descriptive statistics for the sample respondents; Section C identifies potentially useful Census Questions that can be used to measure English language proficiency; Section D describes

Factor Analyses performed on the criterion package questions; Section E presents the results of Regression Analyses with the Census Questions as independent variables and various "ideal" MELPs (e.g., list membership) as the dependent variables; Section F gives the results of Discriminant Analyses designed to classify individuals as LESA or Non-LESA by use of the Census Questions; Section G compares the results of using reverse scales on four of the Census Questions; and Section H discusses a substudy conducted to compare bilingual versus monolingual interviewers. Definitions of many of the variables used in the analyses in this chapter are given in Table V.12 in Section E. The majority of the computations discussed in this chapter were carried out by using the SPSS computer package [5].

B. Description of Sample Members

Tables V.1 to V.10 present descriptive data by site on the sample respondents for the current study. The data presented was obtained from the Household Information Form and Items 2 and 3 of the Census Questionnaire. In particular, Tables V.7 and V.8 show the origin or descent of the children and adults in the sample while Tables V.9 and V.10 give their state, territory or foreign country of birth. Tables V.1 to V.6 present precentage distributions of size of household, availability of telephone, education of head of household, employment status of head of household, occupation of head of household and family income.

(Text continued on page 47)

Table V.1 Size of Household, Percentage Distribution from Item 1 of Household Information Form

Number of persons	Miami		El Paso		San Francisco	NE Arizona
	Child	Adult	Child	Adult	Child	Child
1	0.0	10.6	0.0	2.6	0.0	0.0
2	1.5	31.8	0.9	6.6	0.8	0.0
3	12.5	19.5	6.8	15.9	6.2	2.8
4	33.5	13.2	14.6	12.5	27.2	5.6
5	25.4	12.3	23.8	19.9	21.4	14.6
6	13.4	6.6	20.5	16.2	20.2	20.0
7	7.3	4.0	13.4	11.8	14.4	14.6
8	4.1	0.9	9.6	7.7	5.3	14.0
9	1.5	0.6	3.5	1.8	2.5	7.9
10	0.6	0.3	4.0	1.8	1.2	7.9
11	0.3	0.3	1.9	1.5	0.0	7.9
12	0.0	0.0	0.2	1.1	0.0	2.2
13	0.0	0.0	0.5	0.4	0.4	1.1
14	0.0	0.0	0.0	0.0	0.0	1.7
15	0.0	0.0	0.2	0.0	0.4	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0
Number of households	343	349	425	271	243	178

Table V.2 Availability of Telephone, Percentage Distribution from Item 5 of Household Information Form

Telephone in house	Miami		El Paso		San Francisco	NE Arizona
	Child	Adult	Child	Adult	Child	Child
Yes	85.5	92.8	76.8	85.6	98.4	25.0
No	14.5	7.2	23.2	14.4	1.6	75.0
Total	100.0	100.0	100.0	100.0	100.0	100.0
Number of households	337	335	422	270	243	180

Table V.3 Education of Head of Household, Percentage Distribution from Item 6 of Household Information Form

Highest grade enrolled	Miami		El Paso		San Francisco	NE Arizona
	Child	Adult	Child	Adult	Child	Child
None	0.0	1.0	0.8	4.1	2.6	10.7
Grades K-7	48.6	31.3	36.0	56.9	18.1	28.9
Grade 8	14.2	11.8	7.1	6.9	2.2	10.7
Grades 9-11	14.8	11.2	15.2	11.8	9.2	16.4
Grade 12	11.1	14.7	17.5	12.6	20.7	19.5
College 1-3	4.3	6.7	14.0	6.1	16.3	10.1
College 4	3.7	11.2	5.3	0.8	18.1	2.5
Post graduate	3.4	12.1	4.1	0.8	12.8	1.3
Total	100.0	100.0	100.0	100.0	100.0	100.0
Number of households	325	313	394	246	227	159

Table V.4 Employment Status of Head of Household, Percentage Distribution from Item 7 of Household Information Form

Currently employed	Miami		El Paso		San Francisco	NE Arizona
	Child	Adult	Child	Adult	Child	Child
Yes	72.1	55.5	78.8	80.5	86.9	70.9
No	27.9	44.2	21.2	19.1	13.1	28.5
Don't know	0.0	0.3	0.0	0.4	0.0	0.6
Total	100.0	100.0	100.0	100.0	100.0	100.0
Number of households	344	344	425	272	244	179

Table V.5 Occupation of Head of Household, Percentage Distribution from Item 8 of Household Information Form

Occupation	Miami		El Paso		San Francisco	NE Arizona
	Child	Adult	Child	Adult	Child	Child
Professional & technical	3.1	17.1	13.0	4.4	17.3	12.3
Managers, exc. farm	8.7	8.0	7.9	4.9	11.4	14.0
Sales	1.3	5.1	2.2	0.5	1.5	0.0
Clerical	3.9	8.6	12.1	6.4	10.4	9.7
Craftsmen	31.4	24.6	29.2	25.5	12.9	22.8
Operatives, exc. transport	17.5	14.3	10.2	19.1	4.5	8.8
Transport equip. operative	4.4	1.7	5.4	4.4	0.5	5.3
Laborers, exc. farm	14.8	3.4	6.3	12.3	2.0	9.6
Farmers & managers	0.0	0.0	0.0	0.0	0.0	0.9
Farm laborers & foremen	0.9	0.0	0.3	1.5	0.0	0.0
Service, exc. household	14.0	16.6	10.8	19.1	38.6	16.7
Private household workers	0.0	0.6	0.3	0.5	1.0	0.0
Housewife	0.0	0.0	0.0	0.0	0.0	0.0
Military	0.0	0.0	2.2	1.5	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0
Total number of households	229	175	315	204	202	114

Table V.6 Family Income, Percentage Distribution from Item 9 of Household Information Form

Family Income	Miami		El Paso		San Francisco	NE Arizona
	Child	Adult	Child	Adult	Child	Child
$0- $4,999	32.5	42.7	33.3	44.0	16.9	38.6
$5,000- $9,999	46.0	33.3	32.5	36.4	37.6	30.7
$10,000-$14,999	17.5	16.7	22.0	15.3	25.8	17.1
$15,000-$19,999	2.2	5.3	6.1	3.8	10.8	7.1
$20,000 and over	1.8	2.0	6.1	0.5	8.9	6.4
Total	100.0	100.0	100.0	100.0	100.0	100.0
Number of households	274	300	345	209	213	140

Table V.7 Origin or Descent of Sample Children, Percentage Distribution from Item 2 of Census Questionnaire (Second Hand Reports)

Origin or Descent	Miami	El Paso	San Francisco	NE Arizona
Black	-	1.2	-	-
Central or So. America	4.7	0.2	-	-
Chicano	0.3	0.9	-	-
Chinese	-		50.3	-
Cuban	84.6	-	0.3	-
English	-	1.6	-	0.4
Filipino	0.6	-	30.4	-
French	-	0.2	-	-
German	-	-	-	-
Greek	-	-	-	-
Irish	-	-	-	-
Italian	-	0.2	-	-
Japanese	-	-	12.1	-
Korean	-	-	0.6	-
Latino	-	-	-	-
Mexican	-	28.2	-	0.4
Mexican-American	-	38.6	-	-
Mexicano	0.3	11.1	-	-
Navajo	-	0.5	-	95.7
Negro	0.3	1.4	-	-
(Other) Spanish	1.7	1.4	-	-
Polish	-	-	-	-
Portuguese	-	-	-	-
Puerto Rican	1.7	0.5	-	-
Russian	-	-	-	-
Scottish	-	0.2	-	-
Welsh	-	-	-	-
Vietnamese	-	0.2	-	-
Other	5.5	13.9	5.6	3.6
Don't know	0.3	-	0.6	-
Total	100.0	100.0	100.0	100.0
Number of responses	344	425	322	280

Table V.8 Origin or Descent of Sample Adults, Percentage Distribution from Item 2 of Census Questionnaire (First Hand Reports)

Origin or Descent	Miami	El Paso	San Francisco	NE Arizona
Black	-	-	0.4	-
Central or So. America	0.9	-	-	-
Chicano	0.3	0.8	-	-
Chinese	0.3	-	49.1	-
Cuban	91.7	-	-	-
English	-	-	-	-
Filipino	-	-	33.5	-
French	-	-	-	-
German	-	0.4	-	-
Greek	-	-	-	-
Irish	-	-	0.4	-
Italian	-	0.4	-	-
Japanese	-	0.8	10.9	-
Korean	-	0.8	0.9	3.0
Latino	-	0.8	-	-
Mexican	-	64.0	-	-
Mexican-American	-	5.7	-	-
Mexicano	-	22.3	-	-
Navajo	-	-	-	91.7
Negro	-	-	-	-
(Other)Spanish	4.2	0.8	0.4	-
Polish	-	-	-	-
Portuguese	-	-	0.4	-
Puerto Rican	0.3	0.8	-	-
Russian	-	-	-	-
Scottish	-	-	-	-
Welsh	-	-	-	-
Vietnamese	-	-	-	-
Other	2.4	2.7	3.9	5.4
Don't know	-	-	-	-
Total	100.0	100.0	100.0	100.0
Number of responses	336	264	230	168

Table V.9 State, U.S. Territory, or Foreign Country of Birth for Sample Children, Percentage Distribution from Item 3 of Census Questionnaire (Second Hand Reports)

Where born	Miami	El Paso	San Francisco	NE Arizona
This State	13.5	72.7	45.0	84.6
Different State	8.2	15.0	4.7	15.0
Puerto Rico	1.2	0.2	-	-
Guam	-	0.2	0.3	0.4
Samoa	-	-	0.6	-
China	-	-	21.4	-
Cuba	72.6	-	0.3	-
France	-	-	-	-
Germany	-	0.7	-	-
Greece	-	-	-	-
Italy	-	-	-	-
Japan	-	-	3.4	-
Korea	-	-	0.3	-
Mexico	0.6	10.5	0.3	-
Philippines	-	-	21.1	-
Portugal	-	0.2	-	-
Other	3.8	0 2	2.5	-
Don't know	-	0.2	-	-
Total	100.0	100.0	100.0	100.0
Number of responses	340	421	322	280

Table V.10 State, U.S. Territory, or Foreign Country of Birth for Sample Adults, Percentage Distribution from Item 3 of Census Questionnaire (First Hand Reports)

Where born	Miami	El Paso	San Francisco	NE Arizona
This State	-	7.6	17.3	80.8
Different State	-	8.7	4.4	18.6
Puerto Rico	0.6	0.8	-	-
Guam	0.3	-	-	0.6
Samoa	-	-	0.4	-
China	-	-	38.7	-
Cuba	95.2	-	-	-
France	-	-	-	-
Germany	-	0.4	-	-
Greece	-	-	-	-
Italy	-	0.4	-	-
Japan	-	0.8	4.9	-
Korea	-	0.8	0.9	-
Mexico	-	79.5	-	-
Philippines	-	-	31.1	-
Portugal	-	-	0.4	-
Other	3.9	1.1	1.8	-
Don't know	-	-	-	-
Total	100.0	100.0	100.0	100.0
Number of responses	335	264	225	167

C. Identifying Useful Census Questions

In what follows, the term "first hand" data refers to information collected from the household respondent about himself. Conversely, "second hand" data refers to information collected from the household respondent about another member of the household, not the respondent himself. As described in Chapter IV the term "list membership" refers to a within study site categorization of respondents. In the case of children, the categorization is based upon school administrative decision respecting the ability of each child to function in an English speaking classroom situation. In some sites, the lists involved two categories, in others, three (see Table V.12 in Section E). Similar adult lists were available in only two sites. Criteria used by the schools in making the individual assignments could not be determined in any detail. There is no reason to suppose that the criteria are exactly comparable across sites, even for those sites having the same number of categories. List membership nonetheless reflects the administrative concern of schools for the pupil whose ability in English is likely to influence his academic performance.

The criteria for judging the potential usefulness of census questions were based on the degree of association that exists in the child second hand and the adult first hand sample data sets between the distribution of answers to each census question and list membership. (Note, in this chapter all analysis was carried out using child second hand and adult first hand sample data.) The criterion is concerned with the strength of the association, quite separately from any assessment of its statistical significance. Because of the large size of the sample

data sets, even inconsequential associations were most usually statistically significant at high confidence levels.

The data were arrayed in crosstabulations showing the joint sample distributions of the possible answers to each census question and list membership within each study site. It is convenient in what follows to refer to the answers to the census questions as the rows,

$$r = 1,2,\ldots,R$$

and the list membership as the columns,

$$c = 1,2 \text{ or } 1,2,3$$

in the R by C tables. Two measures of association were computed for each table. These were, Cramér's coefficient of contingency, V ([2], pages 557-560), and the square root of the correlation ratio, η ([2], pages 296-301). For either statistic, the magnitude of the association is measured on the interval [0, 1]. Both statistics can assume the end points of this interval, with the upper limit indicating complete association. If the number of columns,

$$c = 2,$$

then,

$$V = \eta$$

numerically. If

$$c > 2,$$

it is usually the case that

$$\eta \geq V.$$

The computation of these statistics is given in reference [5], page 224 (contingency coefficient) and page 230 (correlation ratio).

It might assist in the interpretation of these statistics to note that for the continuous case,

$$0 \leq \rho^2 \leq \eta^2 \leq 1,$$

where ρ is the usual product moment correlation coefficient. That is, in a regression context, for continuous data, if one were predicting the column means from the rows; then:

(a) $\rho^2 = \eta^2 = 1$, if and only if the rows and columns are in strict linear functional relationship;

(b) $\rho^2 \leq \eta^2 = 1$, if and only if the rows and columns are in strict nonlinear functional relationship;

(c) $\rho^2 = \eta^2 < 1$, if and only if the regression of columns on rows is exactly linear, but there is no exact functional relationship;

(d) $\rho^2 < \eta^2 < 1$ implies there is no exact functional relationship and some nonlinear regression curve is a better fit than the best straight line. ([2], page 297).

Another point of some importance is that η^2 is invariant under permutation of the column array.

The Census Questionnaire used in the study is given in Appendix A. In this section, the census questions are referred to by their number on the questionnaire.

Table V.11 lists the census question numbers which were found to satisfy each of five criteria. The criteria, listed below, establish a minimum value for the association statistics over a minimum number of sites. That is, not only is the magnitude of the association between the census question and list membership of importance, but the question must perform well in a majority of study sites. The criteria are as follows.

Criterion 1: census questions having a value of V and/or $\eta \geq 0.25$ in every site.

Criterion 2: census questions having a value of V and/or $\eta \geq 0.25$ in at least four of the five sites.

Criterion 3: census questions having a value of V and/or $\eta \geq 0.25$ in at least three of the five sites.

Criterion 4: census questions having a value of V and/or $\eta \geq 0.20$ in every site.

Criterion 5: census questions having a value of V and/or $\eta \geq 0.20$ in at least four of the five sites.

Using the results given in Table V.11, RTI and CAL personnel then selected which census questions would be used for further examination. A description of the selected questions after some recoding is given Table V.12.

Table V.11 List of Census Questions Satisfying Screening Criteria.

Criterion 1		Criterion 2		Criterion 3		Criterion 4		Criterion 5	
Adult	Child	Adult	Child	Adult	Child	Adult	Child	Adult	Child
21	4*	21	4*	6	2	6	4*	2	7*
22	12a	22	6	7	4*	21	7	6	6
26*	12d	24	7	21	6	22	12a	21	7
	24	26*	10	22	7	26*	12d	22	10
			12a	24	10		24	24	11
			12b	26*	11			26*	12a
			12c	27	12a			27	12b
			12d	32	12b				12c
			13		12c				12d
			21		12d				13
			24		13				14c
			27		19				21
					21				22
					22				24
					23				27
					24				
					26*				
					27				
					32				

* The question pertains only to a restricted domain.
(e.g. question 4 only applies to immigrants)

D. Factor Analysis

This section describes factor analyses performed on the questions that made up the CP (criterion packages) score. Recall, as described in Chapter IV, that the CP score for children is a combined score from 32 items on the Mat-Sea-Cal Test (MSC) and 15 items on the Oral Communications Test (OCT) while the CP score for adults is a combined score from 16 items on the Adult Production Test, 15 items on the OCT and 10 items on the Adult Comprehension Test (ACT). Before the factor analysis was performed; however, the following data editing was carried out at the request of CAL.

Among the census questions are four questions which are concerned with languages spoken in the household and the language spoken most frequently by the sample individual. Specifically, questions 8 and 9 in combination identify the usual language used by the sample individual. Question 31 identifies the usual language spoken in the household, while question 32 identifies multilingual households. (See Appendix A).

Early in the data analysis activities, following the initial screening of the census questions (Section V.C), the Center for Applied Linguistics requested that the analyses proceed using only a subset of languages reported for these questions. Specifically, new data sets were created which contained only the information for sample individuals;

(a) in Miami and El Paso, having at least one of questions 31, 32, and the 8/9 combination coded Spanish;

(b) in San Francisco, having at least one of questions 31, 32, and 8/9 combination coded Cantonese, Mandarin, Tagalog, Visayam, Ilocano, Japanese, Vietnamese, or Korean;

(c) in Arizona, having at least one of questions 31, 32,

and 8/9 combination coded American Indian or Navaho. The factor analysis (and other analyses described in this report) were conducted on this restricted data set.

The restriction serves to exclude sample individuals whose usual language, as defined above, was coded as German, Italian, French, Polish, Portuguese, English, and Other (included in the sample data set with at least one observation), unless one of the target languages was also spoken by the sample individual or in the household. The frequency of the excluded languages other than English was likely sufficiently small as to fail to influence the results of the factor analysis (or other analyses) regardless of whether they were included or excluded. The exclusion of anglophilic individuals means that the relations quantified in subsequent analyses do not include any points in the relation corresponding to obvious English proficiency. Thus, the scale appropriate for the interpretation of the relations expressed does not extend all the way to complete English dominance. For example, factor score coefficients (and regression coefficients) have no predictive ability with respect to totally English speaking individuals. This fact should be kept in mind throughout this report.

Having edited the data as described above, RTI preceeded to apply factor analysis to the CP questions. The reason for undertaking the factor analysis is that in any testing procedure the results obtained are usually integrated measures of several respondent characteristics. Thus, the questions that make up the CP score are likely to be composite measures of English proficiency, intelligence, possibly socieconomic experience and other characteristics of respondents. It was important

to this study to isolate, so far as was possible, that component of the total score which measured English proficiency. Then the relation between the modified criterion package scores and, for example, the census questions could be interpreted in terms of English proficiency, free of other characteristics which may be common to both the unmodified scores and the questions, and otherwise acting to artificially inflate the magnitude of the relation. Factor analysis provides a procedure for isolating the required component.

Factor analyses were conducted on the matrix of product moment correlations among the 47 criterion package items for children and the 41 criterion package items for adults. The numerical quality of these matrices, given the characteristics of the computing algorithm (SPSS) were too poor to permit anything but the extraction of principal factors (principal components), without iterative communality estimates. The first five factors were extracted, and rotated to orthoganal terminal factors using a quartimax rotation. After some study by CAL personnel of the resulting 47 by 5 factor matrix for children and the 41 by 5 factor matrix for adults and the matrices of factor score coefficients, the decision was reached to define a new MELP determined by the factor scores generated from the coefficients corresponding to the <u>first</u> factor only. In what follows, the term FCTR refers to this factor score. For analysis purposes FCTR was scaled so that it had a mean of zero (over all sites) and a standard deviation of one.

E. Regression Analysis

Having identified census questions that appeared to be related to List membership, RTI then carried out extensive multiple regression analysis using the identified census questions as independent variables and List membership, CP score, CP factor score = FCTR (determined in Section D - above) and the DORP as dependent variables. That is, the following equation was used to investigate the simultaneous relationship between a particular measure of English language proficiency (Y, e.g. CP score) and the identified census questions ($X_1 \ldots X_k$):

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_k X_{ki} + e_i \qquad i = 1,\ldots,n \qquad (V.1)$$

where α, β_1, $\beta_2 \ldots \beta_k$ are unknown parameters to be estimated and e_i is a random error term. Note that equation (V.1) assumes a linear relationship between Y and $X_1,\ldots X_k$. Standard multiple regression analysis was used to estimate the parameters in (V.1). The actual computations were carried out by using the subprogram Regression in the SPSS computer package. Before presenting the results of the regression analysis it should be noted that all of the assumptions underlying regression analysis are not satisfied by the current data. For example, the X variables are certainly not measured without error. However, the purpose of the analysis was only to give an indication of the relationship between the various variables and not to estimate precisely the parameters given in equation (V.1).

In particular, the following multiple regressions were computed for both children and adults:

Dependent Variables (Y)	Independent Variables Children (X)	Independent Variables Adults
(1) CP	When, Speak, Und, Sib, Frnd, Hlang, Years, Birth, Grade, Ped.	When, Speak, Und, Kid Frnd, Hlang, Years, News, Birth, Grade, Incm.
(2) FCTR	Same	Same
(3) LIST	Same	Same
(4) LSTCD	Same	Same
(5) DORP	Same	Same
(6) CP + DORP	Same	Same
(7) FCTR + DORP	Same	Same

Definitions of the various variables and how they were coded are given in Table V.12. The regressions were run by site and over sites and the results are given in Table V.13 for children and Table V.14 for adults at the end of this section. The tables present for each regression computed the standardized regression coefficients, the percent of variation accounted for by the regression (= R^2 which is the square of the correlation coefficient), sample size and the results of tests of significance for the various partial regression coefficients. A standardized regression coefficient (b_j') for independent variable j is defined as

$$b_j' = \frac{b_j s_j}{s_y}$$

where b_j is the partial regression coefficient for variable j and s_j, s_y are the standard deviations for variable j and dependent variable Y, respectively. Thus, standardized regression coefficients are unitless.

(Text continued on page 60)

Table V.12 Definitions of Recoded Variables and Sites Used in Multiple Regression and Discriminate Analyses

	Variable	Description	Code
dependent riables rom Census uestion- ire	When	When came to U.S.?	1 = 1973 or later 2 = Before 1973 or missing 3 = Born in U.S.
	Speak	How well English spoken?	5 = Very well 4 = Well, adequate for most, . adequate 3 = Adequate for few purposes 2 = Just a little, missing 1 = Not at all
	Und	How well English understood?	Same as for SPEAK
	Sib	Language spoken with siblings?	1 = non-English 2 = None, DK or missing 3 = English
	Frnd	Language spoken with best friend?	Same as for SIB
	Hlang	Usual language of household?	Same as for SIB
	Kid	Language spoken with children?	Same as for SIB
	Years	Years of schooling in English?	0 = 0 or missing, otherwise as recorded
	Birth	Year of birth?	As recorded
	Grade	Highest grade of regular school completed?	As recorded
	Ped	Education of household head?	1 = None 2 = Less than grade 8 3 = Less than grade 12 4 = Grade 12 5 = Some college 6 = College graduate 7 = Post graduate
	News	How often read an English-language newspaper?	1 = Often 2 = Occasionally 3 = Not at all, missing
	Incm	Family income	1 = $0-$4,999 2 = $5,000-$9,999 3 = $10,000-$14,999 4 = $15,000-$19,999 5 = $20,000 and over

Table V.12 Continued

	Variable	Description	Code
ependent riables	CP	Total CP score	Children = combined score from 32 items on the MSC and 15 items on the OCT. (Scale 0 to 67) Adult = combined score from 16 items on the APT, 15 items on the OCT and 10 items on the ACT.[1/] (scale 0 to 57)
	FCTR	First factor computed from running a factor analysis on the questions that make up the CP score (see Section D)	Scaled so that it had a mean of zero and a standard deviation of one over all sites
	LIST (Children)	List membership	1,2,3 in sites 1 and 5 1,2 in sites 3,4 and 6
	LSTCD (Children)	List membership[2/]	0,1 in sites 1 and 5 Same as LIST in sites 3,4 and 6
	LIST (Adult)	List membership	1,2,3 in site 1 1,2 in site 3 No lists in sites 4,5 and 6
	LSTCD (Adult)	List membership	0,1 in site 1 Same as LIST in site 3
	DORP	Interviewers direct rating of respondent's understanding and speaking proficiency	5 point scale

1/ A description of the criterion packages (e.g., MSC, OCT) is given in Chapter IV.

2/ Note: LSTCD = LIST when LIST has only two categories (Sites 3,4 and 6)
0 when LIST = 1 or 2 in sites 1 and 5
1 when LIST = 3 in sites 1 and 5

Thus, LSTCD has only two categories in all sites. For discriminate analysis, LSTCD = 0 = LESA.

Table V.12 Continued

Site	Description
Site 1	Miami
Site 3	El Paso
Site 4	San Francisco non-Chinese
Site 5	N.E. Arizona (=Window Rock and Ganado) for dependent variables CP and FCTR; Ganado only for dependent variables LIST and LSTCD *
Site 6	S.F. Chinese

* LIST in Window Rock was considered an unreliable measure of English proficiency.

For children the results of the multiple regression analyses indicated the following:

(i) R^2's for CP are somewhat higher than for FCTR; for example, the R^2 for CP over all sites = .55 while the corresponding R^2 for FCTR = .45.

(ii) As would be expected the R^2's for LIST and LSTCD are smaller than for CP and FCTR (due to the fact that LIST only has values = 1, 2 and 3).

(iii) R^2's for DORP are approximately the same as for FCTR.

(iv) R^2's for standardized (DORP + CP) are approximately the same as for CP alone while R^2's for standardized (DORP + FCTR) are somewhat higher than for FCTR and DORP by themselves.

(v) The CQ variables which appear to be the most important predictors of the various MELP's (i.e., dependent variables) are the Speak, Und, Frnd, Hlang and Years variables. The When and Grade variables do not appear to be important predictors of the MELP's.

For adults the results of the multible regression analyses indicated.

(i) The R^2's for CP and FCTR are approximately the same with a value over all sites of approximately .65. (This is not surprising since for adults the correlation between CP and FCTR is >.96.) This value of .65 is somewhat higher than for children.

(ii) R^2's for LIST and LSTCD are relatively small.

(iii) DORP R^2's are smaller than those for CP and FCTR.

(iv) R^2's for standardized (DORP + CP) and (DORP + FCTR) are approximately the same as for CP and FCTR by themselves.

(v) For CP, FCTR and DORP the most important CQ predictors appear to be Speak, Years and News with Birth and Grade also significant in many cases. For LSTCD the most important predictors are Speak, News and Birth. It is interesting to note for adults that the variables Kid, Frnd and Hlang (i.e. language spoken to various individuals) are not important predictors.

After examining the results of the regression analyses for children, it seemed appropriate to examine the effect of reducing the number of CQ variables used to predict the various MELP's. Accordingly, the following 3 independent variables were defined for children (note that these 3 variables depend upon 6 of the original 10 independent variables).

1. Spund = Speak + Und (Scale = 2-10)
2. Years = same as in Table V.12
3. Usage = Sib' + Hlang' + Frnd' (Scale = 0-3)

where Sib' = 0 if Sib = 1 or 2
= 1 if Sib = 3

Hlang' = 0 if Hlang = 1 or 2
= 1 if Hlang = 3, and

Frnd' = 0 if Frnd = 1 or 2
= 1 if Frnd = 3.

Thus, Spund is a combined measure of how well English is spoken and under-

stood and Usage is a combined measure of the language used when speaking to various individuals.

Using the above three independent variables, multiple regressions were computed for children for the dependent variables CP, FCTR, LIST, LSTCD and DORP. The results which are given in Table V.15 indicate the following (note Table V.15 gives the R^2's for the 3 independent variable model as well as the 10 independent variable model):

(i) In general the R^2's over sites are approximately the same for the 3 independent variable model as compared with the 10 independent variable model. This indicates that the 3 variables Spund, Usage and Years can predict CP, FCTR, etc. about as well as all 10 of the original independent variables.

(ii) For the various sites, the 3 variable model appears to predict as well as the 10 variable model in Sites 1 and 3 but not quite as well in Sites 4, 5 and 6.

(iii) The tests of significance of the partial regression coefficients for Spund, Usage and Years are almost always significant for each of the dependent variables. Also, it appears that Years is a much more important predictor for CP and FCTR than it is for LIST and LSTCD.

(Text continued on page 70)

Table V.13 Summary of Regression Analyses for Children With 10 Census Questions as Independent Variables and Several Different Dependent Variables, Body of Table Gives Standardized Regression Coefficients, Regressions Run by Sites and Over Sites.

Independent Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. non-Chinese)	Site 5 (N.E. Arizona)	Site 6 (S.F. Chinese)	Over Sites	
Dependent Variable = CP							
When	-.006	.057	.037	-.105*	.047	-.002	
Speak	.139*	.200**	.044	.247**	.151	.208**	
Und	.251**	.112	.323*	-.047	.128	.128**	
Sib	.027	.006	.092	-.028	.057	.058*	
Frnd	.118**	.274**	.053	.193**	.113	.200**	
Hlang	-.032	.196**	.033	.184*	-.060	.073**	
Years	.317**	.141**	.091	-.092	.150	.219**	
Birth	-.084	-.111*	-.241	-.032	-.352**	-.143**	
Grade	.114	-.005	.159	.471**	.099	.054	
Ped	.015	.088*	.146	.106*	-.033	.066**	
N[1/]	317	364	133	260	146	1220	
R^2	.57	.59	.44	.50	.56	.55	
Dependent Variable = FCTR							
When	-.015	.135**	-.053	-.063	.028	.016	
Speak	.138	.199**	-.147	.282**	.031	.194**	
Und	.230**	.105	.490**	-.067	.217	.129**	
Sib	.045	.079	.144	-.039	.077	.078**	
Frnd	.131**	.238**	-.121	.190**	.209**	.204**	
Hlang	-.052	.105	.115	.185*	-.120	.049	
Years	.281**	.131**	.201	-.123	.090	.198**	
Birth	-.068	-.074	-.070	.044	-.293*	-.077*	
Grade	.036	-.010	.095	.452**	.075	.020	
Ped	-.018	.066	.061	.046	.039	.053*	
N	317	364	133	260	146	1220	
R^2	.45	.53	.37	.37	.43	.45	
Dependent Variable = LIST (Ganado)							
When	.063	.077*	.309**	.080	.092		
Speak	.234**	.141*	-.167	.306*	.318*	Consistent definition of LIST not available over sites	
Und	.107	.148*	.264	-.246	-.093		
Sib	.067	.166**	.093	.009	.059		
Frnd	.057	-.004	.066	.231**	.107		
Hlang	.014	.372**	.078	.143	.127		
Years	.298**	.009	.089	.004	.162		
Birth	-.023	.043	-.220	.183	-.028		
Grade	.037	.032	-.170	.368*	-.071		
Ped	-.023	.020	.089	.144	.131		
N	317	384	133	138	146		
R^2	.40	.59	.27	.30	.31		

Table V.13 Continued

Dependent Variable = LSTCD

Independent Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. non-Chinese)	Site 5 (Ganado)	Site 6 (S.F. Chinese)	Over Sites
When	.059			.041		.059*
Speak	.259**			.374*		.209**
Und	.012			-.212		.059
Sib	.018	Same	Same	-.041	Same	.116**
Frnd	.052	as	as	.236*	as	.089**
Hlang	.034	List	List	.060	List	.195**
Years	.299**			.040		.088*
Birth	.032			.222		.041
Grade	.069			.141		.020
Ped	.027			.187*		.063*
N	317			138		1098
R^2	.31			.22		.35

Dependent Variable = DORP

When	.130*	-.039				.018
Speak	.341**	.100				.220**
Und	.058	.158*	Insuffi-	Insuffi-	Insuffi-	.125**
Sib	.007	.064	cient Data	cient Data	cient Data	.043
Frnd	.084	.225**				.144**
Hlang	.050	.283**				.151**
Years	.270**	.042				.183**
Birth	-.023	-.078				-.073
Grade	.061	-.069				-.019
Ped	.022	.104*				.062*
N	307	306				855
R^2	.48	.57				.42

	Dependent Variable = Standardized (CP + DORP)			Dependent Variable = Standardized (FCTR + DORP)		
	Site 1	Site 3	Over Sites	Site 1	Site 3	Over Sites
When	.075	-.003	.002	.075	.029	.014
Speak	.269**	.148*	.214**	.274**	.154*	.217**
Und	.151*	.189**	.166**	.135	.186**	.164**
Sib	.027	.054	.051	.035	.102	.065*
Frnd	.104*	.250**	.176**	.115*	.241**	.183**
Hlang	.022	.259	.145**	.020	.206**	.134**
Years	.307**	.108*	.236**	.292**	.099*	.225**
Birth	-.063	-.107*	-.118**	-.064	-.085	-.092*
Grade	.101	-.055	.009	.054	-.055	-.013
Ped	.031	.092*	.072**	.014	.080	.063*
N	307	306	855	307	306	855
R^2	.59	.68	.56	.53	.67	.54

1/ N = Sample size, R^2 = percent of variation accounted for by the regression.
* = Partial regression coefficient significant at .05 level.
** = Partial regression coefficient significant at .01 level.

Standardized regression coefficient $b_i' = b_i \frac{s_i}{s_y}$ where b_i is the partial regression coefficient.

Table V.14 Summary of Regression Analyses for Adults With 11 Census Questions as Independent Variables and Several Different Dependent Variables, Body of Table Gives Standarized Regression Coefficients, Regressions Run by Site and Over Sites.

Independent Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. non-Chinese)	Site 5 N.E. (Arizona)	Site 6 (S.F. Chinese)	Over Sites
			Dependent Variable = CP			
When	.129**	.151*	.188*	-.042	.073	.076**
Speak	.299**	.220	.205	.132	.476**	.266**
Und	.207**	.218	.021	.042	-.128	.108*
Kid	-.021	.049	-.079	.135*	-.026	.019
Frnd	.017	-.069	.026	.040	.104	.034
Hlang	-.066	-.099	.049	.070	.023	-.014
Years	.032	.270**	.089	.203	.225*	.204**
News	-.126*	-.194**	-.210**	-.266**	-.029	-.147**
Birth	.096*	-.014	.297**	.079	.142*	.112**
Grade	.228**	.048	.135	.162*	.201**	.163**
Incm	.007	.148*	.259**	-.097	.016	.047*
N[1/]	272	202	116	214	111	915
R^2	.50	.40	.56	.57	.70	.67
			Dependent Variable = FCTR			
When	.113*	-.155*	.159	-.030	.065	.076**
Speak	.284**	.203	.143	.108	.461**	.246**
Und	.211**	.258**	.067	.073	-.123	.134**
Kid	-.030	.027	-.085	.143*	-.008	.017
Frnd	-.002	-.065	.067	.047	.101	.034
Hlang	-.092*	-.106	.028	.064	-.024	-.026
Years	.019	.267**	.120	.149	.306**	.209**
News	-.131*	-.214**	-.216**	-.273**	.009	-.143**
Birth	.092	-.010	.283**	.071	.115	.105**
Grade	.227**	.036	.149	.174*	.167*	.155**
Incm	.012	.134*	.208**	-.103	.030	.041
N	272	202	116	214	111	915
R^2	.47	.42	.53	.53	.67	.65
			Dependent Variable = LIST			
When	-.046	.074				
Speak	.398**	.158				
Und	-.144	-.125	No List Available	No List Available	No List Available	Consistent Definition of List Not Available Over Sites
Kid	-.020	.053				
Frnd	.016	-.086				
Hlang	.126*	-.008				
Years	.064	.010				
News	-.126*	-.179*				
Birth	.134*	.048				
Grade	.234**	-.010				
Incm	-.005	-.097				
N	272	202				
R^2	.29	.05				

Table V.14 Continued

Dependent Variable = LSTCD

Independent Variables	Site 1	Site 3	Site 4	Site 5	Site 6	Over Sites
When	-.008					.092*
Speak	.310**					.227**
Und	-.082					-.108
Kid	.002	Same	No List	No List	No List	-.021
Frnd	-.004	as	Avail-	Avail-	Avail-	-.026
Hlang	.082	List	able	able	able	.042
Years	.058					.025
News	-.106					-.173**
Birth	.132*					.210**
Grade	.228**					.060
Incm	.006					-.045
N	272					474
R^2	.23					.12

Dependent Variable = DORP

When	.099	-.185*				.076*
Speak	.259**	.097				.232**
Und	.126	.276				.063
Kid	-.002	.164	Insuffi-	Insuffi-	Insuffi-	.020
Frnd	.121*	-.158	cent Data	cent Data	cent Data	.050
Hlang	-.045	.043				.042
Years	.128	.180*				.262**
News	-.153**	-.161*				-.151**
Birth	.023	-.033				.072*
Grade	.128*	.001				.097**
Incm	.013	.141				.035
N	262	153				605
R^2	.40	.29				.59

Table V.14 Continued

	Dependent Variable = Standardized (CP + DORP)			Dependent Variable = Standardized (FCTR + DORP)		
	Site 1	Site 3	Over Sites	Site 1	Site 3	Over Sites
When	.116*	-.217**	.079**	.107*	-.225**	.080**
Speak	.297**	.247	.254**	.288**	.222	.243**
Und	.174*	.186	.076	.179*	.231	.093
Kid	-.008	.140	.029	-.013	.128	.025
Frnd	.075	-.163	.030	.066	-.166	.029
Hlang	-.061	-.039	.007	-.075	-.045	.003
Years	.083	.251**	.288**	.075	.251**	.287**
News	-.153**	-.210**	-.153**	-.155**	-.229**	-.155**
Birth	.067	-.056	.087**	.063	-.058	.084**
Grade	.203**	.060	.144**	.203**	.060	.141**
Incm	.012	.173*	.045	.016	.167*	.040
N	262	153	605	262	153	605
R^2	.51	.39	.69	.49	.41	.68

1/ N = Sample size, R^2 = percent of variation accounted for by the regression.
* = Partial regression coefficient significant at .05 level.
** = Partial regression coefficient significant at .01 level.

Standardized regression coefficient $b_i' = b_i \frac{s_i}{s_y}$ where b_i is the partial regression coefficient.

Table V.15 Summary of Regression Analyses for Children with Independent Variables Spund, Usage and Years and Several Different Dependent Variables, Body of Table Gives Standardized Regression Coefficients, Regressions Run by Sites and Over Sites.

Independent Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. non-Chinese)	Site 5 (N.E. Arizona)	Site 6 (S.F. Chinese)	Over Sites
Dependent Variable = CP						
Years	.411**	.184**	.393**	.343**	.431**	.348**
Spund	.375**	.328**	.384**	.202**	.301**	.344**
Usage	.102*	.432**	.098	.309**	.141	.254**
N[1/]	317	364	133	260	146	1220
R^2	.53(.57)	.56(.59)	.39(.44)	.40(.50)	.47(.56)	.52(.55)
Dependent Variable = FCTR						
Years	.330**	.156**	.341**	.249**	.301**	.259**
Spund	.353	.332**	.317**	.200**	.280**	.328**
Usage	.124*	.392**	.123	.296**	.183*	.271**
N	317	364	133	260	146	1220
R^2	.43(.45)	.50(.53)	.30(.37)	.30(.37)	.35(.43)	.43(.45)
Dependent Variable = LIST						
				(Ganado)		
Years	.334**	-.004	.070	.217*	.127	Consistent Definition of LIST not Available Over Sites
Spund	.324**	.321**	.156	.045	.239**	
Usage	.117*	.473**	.332**	.354**	.283**	
N	317	364	133	138	146	
R^2	.40(.40)	.55(.59)	.19(.27)	.21(.30)	.27(.31)	
Dependent Variable = LSTCD						
				(Ganado)		
Years	.325**	Same	Same	.030	Same	.057*
Spund	.270**	as	as	.141	as	.274**
Usage	.066	List	List	.269**	List	.357**
N	317			138		1098
R^2	.30(.31)			.13(.22)		.34(.35)

Table V.15 Continued

Dependent Variable = DORP

Independent Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. non-Chinese)	Site 5 N.E. (Arizona)	Site 6 (S.F. Chinese)	Over Sites
Years	.312	.039	Insuffi-	Insuffi-	Insuffi-	.209
Spund	.393	.273	cient Data	cient Data	cient Data	.345
Usage	.127	.509				.285
N	307	306				855
R^2	.46(.48)	.54(.57)				.41(.42)

1/ N = Sample size, R^2 = percent of variation accounted for by the regression.
* = Partial regression coefficient significant at .05 level.
** = Partial regression coefficient significant at .01 level.

Standardized regression coefficient $b_i' = b_i \frac{s_i}{s_y}$ where b_i is the partial regression coefficient.

For R^2: the number in ( ) gives the R^2 when 10 independent variables were used to compute the regression.

F. Discriminant Analysis

As described previously, one of the major purposes of the present study was to develop a method of using the Census - Questions to determine whether or not an individual had limited English speaking ability (i.e., LESA or Non-LESA). In Sections C and E of the chapter we have selected subsets of the Census Questions to use as LESA predictors. One commonly used statistical method to classify individuals into various populations is discriminant analysis which forms a linear combination of the discriminating variables (e.g., the Census Questions) as follows:

$$S = d_0 + d_1X_1 + \ldots + d_kX_k \qquad (V.2)$$

where the d's are unknown parameters to be estimated; $X_1,\ldots,X_k$ are the discriminating variables and S is a classification score.

Equation (V.2) can be utilized to classify individuals into two groups (e.g., LESA or Non-LESA) depending on the value of their classification score. That is for the i^{th} individual if $S_i >$ constant then individual i would be classified as LESA. The linear function in equation (V.2) is called a discriminant function (D.F.). The unknown parameters $d_1,\ldots,d_k$ in equation (V.2) are estimated from sample data for the two groups of individuals so that the discriminant function maximizes the ratio of the distance between the two groups to the within groups spread (e.g., see [4] or [7]).

Accordingly, for the present study, RTI performed discriminant analyses on the sample field test data in order to determine linear functions (D.F.s) based on the Census Questions that would classify individuals as LESA or Non-LESA.

In addition to estimating the discriminate functions, RTI and CAL were naturally interested in what percent of the time these D.F.s would correctly classify individuals as LESA or Non-LESA. Unfortunately, in the present case it was not possible to obtain unbiased estimates of these percentages. However, one commonly used procedure for indicating how well a discriminant function classifies individuals is to (a) estimate the D.F. using sample data from the two groups (i.e., LESA and Non-LESA) (b) classify the same sample data into LESA or Non-LESA groups using the estimated function (c) compare the actual LESA/Non-LESA classification with the classification by the estimated discriminant. The above procedure which is biased (i.e., the procedure underestimates the probability of misclassification, see [3]) is the method used by RTI throughout the present section to examine "how well" the estimated linear discriminants are performing. The necessary computations for estimating the D.F.s and then classifying individuals as LESA or Non-LESA based on these D.F.s were carried out by using the subprogram Discriminant in the SPSS computer package.

In this section, subsection F.1 presents discriminant analyses for children, subsection F.2 investigates discriminant functions to be used over all sites, subsection F.3 discusses a discrete discriminating procedure, subsection F.4 examines estimation of the percent LESA in a sample population and subsection F.5 presents discriminant analyses for adults.

F.1 Discriminant Functions for Children

In particular, RTI first carried out discriminant analyses on the sample field test data for children using the same ten Census Questions as in Section E (Regression Analysis) as the discriminating variables (see Table V.13). In order to perform these analyses, it was first necessary to partition the children into LESA and Non-LESA groups. This was

done by using the variables LSTCD, FCTR and CP (Note in discriminant analysis, it is assumed for the sample field test data that membership in the two groups to be classified is known which was not the case for the present data since an "ideal MELP" has not been defined; thus, the use of three different MELPs.)

Defining the LESA and Non-LESA groups by LSTCD was straightforward since this variable only had two values in every site (see Table V.12). However, for FCTR and CP which are continuous variables it was necessary to define all individuals with a FCTR score (or CP score) less than a constant as LESA and the remaining individuals as Non-LESA. After discussions with CAL staff and some preliminary computer runs, the constant for FCTR was initially set at .2 and for CP at 50. RTI then computed several discriminate analyses using these constants and a few of the results are given in Appendix B. After examining these analyses, CAL decided to run additional analyses with the constant for FCTR set so that the number of LESAs defined by FCTR approximately equaled the number of LESAs defined by LSTCD in each site. Thus, it was necessary for the constant for FCTR to be set at different values in each site. In addition, since the preliminary results for CP were quite similar to those of FCTR no additional analyses were done with CP. Accordingly, in the remainder of this subsection no CP analysis is given and the definition of LESA by FCTR changes for each site.[1/]

The results of using discriminant analysis to classify the field test data as LESA and Non-LESA using 10 CQs as discriminators and LSTCD and FCTR for defining LESA and Non-LESA are given in Table V.16.

1/ Note, since the definition of LESA by FCTR is somewhat arbitrary, this definition was not forced to be the same for all analyses in section F. Thus, throughout this section, the definition of LESA by FCTR is indicated whenever the results of a discriminate analysis are presented.

The body of the table gives (i) the number of children who are LESA and Non-LESA as defined by LSTCD and FCTR and (ii) the number of children who are predicted to be LESA and Non-LESA by the discriminant function based on the 10 CQ variables. In addition, the table gives the percent of LESAs classified as Non-LESA (= α_1) by the D.F., the percent of Non-LESAs classified as LESA (= α_2) by the D.F.,the overall percent of individuals classified correctly (= % correct), the estimated percent LESA in the sample using the D.F., and the actual percent LESA as defined by LSTCD (or FCTR). It should be noted here that a different D.F. was computed for each site and over sites and for each definition of LESA (= LSTCD and FCTR). Thus, Table V.16 gives the results for 12 different D.F.s.

Table V.16 indicates that over sites between 75 and 80% of the children are classified the same by the D.F. and LSTCD or FCTR and that the LESA estimate using the D.F. is approximately 55% as compared to the value for LSTCD and FCTR of 58%. By site, the best group prediction appears to be in El Paso (87.1% for LSTCD and 84.9% for FCTR) while Site 5 and Site 4, respectively, have the lowest percent correctly classified by the D.F. for LSTCD (70.3%) and FCTR (70.7%). In most sites the estimated % LESA by the D.F. is too low, as compared to LSTCD and FCTR, particularily in Site 5.

RTI also computed D.F.s for children based only on the three variables Spund, Usage and Years due to the results of the multiple regression analyses described in Section E and; in addition, due to an investigation of the relative sizes of the weights (the d_i's in equation (V.2)) that the 10 CQs received in the 10 variable D.F.s. The results of these computations are given in Table V.17 and V.18. Table V.17 gives the same results for Spund, Usage and Years as were given earlier in Table V.16 for the 10 CQs and; in addition, the table presents a few

(Text continued on page 79)

Table V.16 Results of Using Discriminant Analysis to Predict Whether A Child is LESA or Non-LESA, Discriminant Function Based on 10 Census Questions, LESA Defined by LSTCD and FCTR[1/] [2/]

LSTCD Used to Define LESA

LSTCD Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	162	48	176	15	34	19	66	29	69	24
Non-LESA	23	84	32	141	17	63	12	31	12	41
α_1 [3/]	22.9		7.9		35.8		30.5		25.8	
α_2	21.5		18.5		21.3		27.9		22.6	
% Correct	77.6		87.1		72.9		70.3		75.3	
Est % LESA	58.4		57.1		38.3		56.5		55.5	
Actual % LESA	66.2		52.5		39.8		68.8		63.7	

Over Sites

LSTCD Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	497	145
Non-LESA	111	345
α_1	22.6	
α_2	24.3	
% Correct	76.7	
Est % LESA	55.4	
Actual % LESA	58.5	

Table V.16 Continued

FCTR Used to Define LESA[4/]

FCTR Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	156	53	166	24	36	17	73	21	69	23
Non-LESA	27	81	31	143	22	58	8	36	16	38
α_1	25.4		12.6		32.1		22.3		25.0	
α_2	25.0		17.8		27.5		18.2		29.6	
% Correct	74.8		84.9		70.7		79.0		73.3	
Est % Lesa	57.7		54.1		43.6		58.7		58.2	
Actual % LESA	65.9		52.2		39.8		68.1		63.0	

Over Sites

FCTR Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	504	136
Non-LESA	108	350
α_1	21.3	
α_2	23.6	
% Correct	77.8	
Est % LESA	55.7	
Actual % LESA	58.3	

1/ 10 Census Questions = When, Speak, Und, Sib, Frnd, Hlang, Years, Birth, Grade and Ped.

2/ Discriminant Function computed for each site.

3/ α_1 = Percent of LESAs classified as Non-LESA; α_2 = Percent of Non-LESAs classified as LESA; % correct = overall percent classified correctly.
Est % LESA = estimated percent LESA = 100 x (Number predicted as LESA)/(Total number of children)
Actual % LESA = percent LESAs as given by LSTCD (or FCTR).

4/ Site 1: FCTR ≤.45 = LESA; Site 3: FCTR ≤.18 = LESA;
Site 4: FCTR ≤.54 = LESA; Site 5: FCTR ≤.63 = LESA;
Site 6: FCTR ≤.41 = LESA; Over sites FCTR ≤.43 = LESA

Table V.17 Results of Using Discriminant Analysis to Predict Whether A Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, LESA Defined by LSTCD and FCTR[1/] [4/]

LSTCD Used to Define LESA

LSTCD Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	160	50	161	30	32	21	65	30	69	24
Non-LESA	27	80	27	146	23	57	16	27	16	37
α_1 [2/]	23.8 (22.9)		15.7 (7.9)		39.6 (35.8)		31.6 (30.5)		25.8 (25.8)	
α_2	25.2		15.6		28.8		37.2		30.2	
% Correct	75.7 (77.6)		84.3 (87.1)		66.9 (72.9)		66.7 (70.3)		72.6 (75.3)	
Est LESA	59.0 (58.4)		51.6 (57.1)		41.4 (38.3)		58.7 (56.5)		58.2 (55.5)	
Actual LESA	66.2		52.5		39.8		68.8		63.7	

Over Sites

LSTCD Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	488	154
Non-LESA	108	348
α_1	24.0 (22.6)	
α_2	23.7	
% Correct	76.1 (76.7)	
EST LESA	54.3 (55.4)	
Actual LESA	58.5	

Table V.17 Continued

FCTR[3] Used to Define LESA

FCTR Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	156	53	157	33	36	17	64	30	58	34
Non-LESA	26	82	31	143	25	55	14	30	18	36
α_1	25.4 (25.4)		17.4 (12.6)		32.1 (32.1)		31.9 (22.3)		37.0 (25.0)	
α_2	24.1		17.8		31.3		31.8		33.3	
% Correct	75.1 (74.8)		82.4 (84.9)		68.4 (70.7)		68.1 (79.0)		64.4 (73.3)	
EST LESA	57.4 (57.7)		51.6 (54.1)		45.9 (43.6)		56.5 (58.7)		52.1 (58.2)	
Actual LESA	65.9		52.2		39.8		68.1		63.0	

Over Sites

FCTR Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	471	169
Non-LESA	99	359
α_1	26.4 (21.3)	
α_2	21.6	
% Correct	75.6 (77.8)	
EST LESA	51.9 (55.7)	
Actual LESA	58.3	

1/ Discriminant Function computed for each site.

2/ α_1 = Percent of LESAs classified as Non-LESA; α_2 = Percent of Non-LESAs classified as LESA; % correct = overall percent classified correctly.
Est % LESA = estimated percent LESA = 100 x (Number predicted as LESA)/(Total number of children)
Actual % LESA = percent LESAs as given by LSTCD (or FCTR).

3/ Site 1: FCTR ≤.45 = LESA; Site 3: FCTR ≤.18 = LESA
Site 4: FCTR ≤.54 = LESA; Site 5: FCTR ≤.63 = LESA
Site 6: FCTR ≤.41 = LESA; Over sites FCTR ≤.43 = LESA

4/ Numbers in ( ) give results when 10 CQ variables were used to compute the discriminant function (see Table V.16)

Table V.18 Standardized and Unstandardized Discriminant Function Coefficients for Children for Spund, Usage and Years, LESA Group Defined by LSTCD and FCTR

Discriminating Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. Non-Chinese)	Site 5 (Ganado)	Site 6 (S.F. Chinese)	Over Sites
	LSTCD Used to Define LESA					
	Standardized Coefficients[1]					
Spund	-.49	-.43	-.36	-.39	-.46	-.47
Usage	-.12	-.64	-.76	-.74	-.54	-.62
Years	-.59	.01	-.16	-.08	-.24	-.10
	Unstandardized Coefficients					
Spund	-.22	-.18	-.21	-.17	-.21	-.21
Usage	-.15	-.50	-.82	-.64	-.53	-.54
Years	-.34	.01	-.10	-.04	-.18	-.06
Constant	2.53	2.04	3.88	2.09	2.62	2.36

	Site 1	Site 3	Site 4	Site 5	Site 6	Over Sites
	FCTR[2] Used to Define LESA					
	Standardized Coefficients					
Spund	-.51	-.44	.37	-.39	-.38	-.44
Usage	-.17	-.61	.40	-.54	-.28	-.47
Years	-.54	-.11	.70	-.42	-.62	-.42
	Unstandardized Coefficients					
Spund	-.23	-.19	.21	-.17	-.18	-.19
Usage	-.21	-.48	.42	-.46	-.27	-.41
Years	-.31	-.13	.42	-.21	-.46	-.26
Constant	2.57	2.25	-4.00	2.53	2.73	2.56

[1] Standardized coefficients indicate the relative contribution of each variable to the discriminant function.
Unstandardized coefficients are multiplied by the raw values of Spund, Usage and Years to obtain a discriminant score which is used to classify a child as LESA or Non-LESA.

[2] Site 1: FCTR ≤.45 = LESA; Site 3: FCTR ≤.18 = LESA;
Site 4: FCTR ≤.54 = LESA; Site 5: FCTR ≤.63 = LESA;
Site 6: FCTR ≤.41 = LESA; Over Sites FCTR ≤.43 = LESA.

of the 10 CQs results in ( ) for comparison purposes. Table V.17 indicates that the D.F.s based on the 3 variables do almost as well in correctly classifying children as do the D.F.s based on 10 variables. For example, over sites 76.1% of the children are classified the same by the 3 variable D.F. as by LSTCD while the corresponding percentage for the 10 variable discriminant is 76.7%. By site, the greatest decreases in the percent correctly classified for the 3 versus the 10 variable D.F.s are in Site 4 for LSTCD (66.9% versus 72.9%) and Site 5 for FCTR (68.1% versus 79.0%). To summarize, the D.F.s based on the three variables Spund, Usage and Years appear to classify the field test data as LESA or Non-LESA (as defined by LSTCD and FCTR) almost as well as the D.F.s based on the 10 CQ variables. Recall that the multiple regression analyses in Section E indicated a similar result.

Table V.18 presents standardized and unstandarized discriminant function coefficients for the 3 variable case (i.e., the d_i's in equation (V.2)). The magnitude of the standardized coefficients indicates the relative importance of each of the three variables in classifying children in the field test as LESA or Non-LESA. Thus, when LSTCD is used to define LESA it appears that Years is not as important a discriminator as Spund and Usage (e.g., over sites the standarized coefficient for Years has a magnitude of .10 while the corresponding magnitudes for Spund and Usage are .47 and .62 respectively). However, when FCTR is used to define LESA then over sites all 3 variables are approximately equally important as discriminators.

F.2 Overall Discriminate Function

The discriminate analyses for the field data for children given in subsection F.1 computed a different discriminate function for

each site in the field test. However, in practice it may be desirable to have only one D.F., i.e., a D.F. which can be used for any site or ethnic group. To investigate this possibility, RTI used the 3 variable D.F. computed over sites to classify individuals as LESA or Non-LESA within each site; and then compared the results to the classification obtained by using the within site 3 variable D.F.s. The results of these computations are given in Table V.19. Note that the results for FCTR in the table are based upon LESA being defined by FCTR $\leq$ .2 in all sites. Thus, a consistent definition was used over sites when LESA was defined by FCTR unlike Tables V.16 and V.17 where the cutting point on FCTR was allowed to vary by site. The over site D.F.s used to obtain the results in Table V.19 are given in subsection F.1 and Appendix B.

Table V.19 indicates that for both LSTCD and FCTR the percent correctly classified is about the same using the over site D.F. as compared to using the within site D.F.s. However, the estimated magnitude of the errors of misclassification, α_1 and α_2, are quite different in some sites for the two types of D.F.s. For example, in Site 4 for LSTCD, α_1 = 64.2% and α_2 = 8.8% for the overall D.F. while the corresponding percentages for the within site D.F. are 39.6 and 28.8%. The implications of large differences in α_1 and α_2 for the two D.F.s is that the estimate of the % LESA for the two D.F.s may also be relatively large. Thus, the Site 4 estimates of percent LESA are 19.5% for the overall D.F. versus 41.4% for the within site D.F. In general, it is usually the case that the values of α_1 and α_2 are more nearly equal for the within site D.F. versus the over site D.F. This result may or may not produce a more accurate estimate of the % LESA for the within site D.F. depending on the actual percent LESA in the population being sampled. Subsection F.4 below discusses estimation

Table V.19 Results of Using a Discriminant Function Estimated Over All Sites to Predict Whether a Child is LESA or Non-LESA Within Each Site, Discriminant Function Based on Spund, Usage and Years, LESA Defined by LSTCD and FCTR[1/3/]

LSTCD Used to Define LESA

LSTCD Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	170	40	165	26	19	34	66	29	68	25
Non-LESA	41	66	28	145	7	73	15	28	17	36
α_1	19.0 (23.8)		13.6 (15.7)		64.2 (39.6)		30.5 (31.6)		26.9 (25.8)	
α_2	38.3 (25.2)		16.2 (15.6)		8.8 (28.8)		34.9 (37.2)		32.1 (30.2)	
% Correct	74.4 (75.7)		85.2 (84.3)		69.2 (66.9)		68.1 (66.7)		71.2 (72.6)	
Est % LESA	66.6 (59.0)		53.0 (51.6)		19.5 (41.4)		58.7 (58.7)		58.2 (58.2)	
Actual % LESA	66.2		52.5		39.8		68.8		63.7	

FCTR Used to Define LESA[2/]

FCTR Group	Predicted Group LESA	Predicted Group Non-LESA	Predicted Group LESA	Predicted Group Non-LESA	Predicted Group LESA	Predicted Group Non-LESA	(N.E. Arizona) Predicted Group LESA	(N.E. Arizona) Predicted Group Non-LESA	Predicted Group LESA	Predicted Group Non-LESA
LESA	128	27	161	35	10	12	47	32	45	13
Non-LESA	55	107	26	142	12	99	24	157	30	58
α_1	17.4 (21.3)		17.9 (18.4)		54.5 (22.7)		40.5 (25.3)		22.4 (31.0)	
α_2	34.0 (25.3)		15.5 (16.7)		10.8 (22.5)		13.3 (25.4)		34.1 (26.1)	
% Correct	74.1 (76.7)		83.2 (82.4)		82.0 (77.4)		78.5 (74.6)		70.6 (71.9)	
Est % LESA	57.7 (51.4)		51.4 (51.6)		16.5 (31.6)		27.3 (40.4)		51.4 (43.2)	
Actual % LESA	48.9		53.8		16.5		30.4		39.7	

1/ Definitions of α_1, α_2, % correct, Est % LESA and actual % LESA given in Table V.16

2/ FCTR ≤.2 = LESA in all sites.

3/ Numbers in ( ) give results when discriminant functions were computed for each site (see Tables V.17 and B-2)

of the % LESA in more detail. Table V.19 also indicates, except for Site 4 for LSTCD and FCTR and Site 5 for FCTR, that the estimated percent LESAs in each site are relatively close for the overall D.F. versus the within site D.F.s. The results of using an over site D.F. to classify adults as LESA/Non-LESA within each site are discussed in subsection F.5 (see Table V.28).

F.3. A Discrete Discriminating Rule

Classical discriminate analysis assumes that the discriminating variables are continuous. Obviously this is not true in the present case where the discriminating variables are the CQs which are discrete variables (see Table V.12). Thus, the usual assumptions underlying the discriminate analyses in subsections F.1 and F.2 do not hold. To examine the effect of not satisfying this continuous variable assumption, RTI investigated a discriminating procedure given by Cochran and Hopkins [1] which is designed to handle the classification problem when the discriminating variables are discrete. The procedure (denoted here by the C procedure) is designed to minimize the expected frequency of misclassification. In general, for the present case, the C procedure determines from the field test data, the probability of being LESA and Non-LESA for each cell of a table whose cells are defined by the levels of the discriminating variables. For example, if there were only 2 discriminating variables (V1 and V2) each with 2 levels (L1 and L2) then the probabilities would be estimated for the following 4 cell table:

Value of Discriminating Variables		Cell Probabilities for	Cell Probabilities for
V1	V2	LESA	Non-LESA
L1	L1	P_1	P'_1
L1	L2	P_2	P'_2
L2	L1	P_3	P'_3
L2	L2	P_4	P'_4
Total		1.00	1.00

where P_1 = (number of LESAs in cell V1 = L1, V2 = L1)/(Total number of LESAs in all cells)

P_1' = (number of Non-LESAs in cell V1 = L1, V2 = L1)/(Total number of Non-LESAs in all cells)

etc., for P_2, P_2'...

(note in the present case LESA or Non-LESA are determined by LSTCD or FCTR). Having estimated the probabilities for each cell of the table from the field test, the C procedure then classifies the individuals in a sample population in cell i as LESA if

$$\pi_1 P_i > \pi_2 P_i' \qquad \text{(V.3)}$$

where π_1 is the frequency of LESAs and π_2 is the frequency of Non-LESAs in the sample population. Thus, the C procedure requires an initial estimate or guess of the value of π_1. In practice, since we are dealing with discrete variables, the accuracy of the estimate of π_1 may not be critical because the same classification rule is optimum over a range of values of π_1. Note, however, that the classification rule given by the C procedure will change depending on the values of π_1 and π_2. That is, the classification of the cells in the table defined by the discriminating variables will change depending on the values of π_1, and π_2 (For the classical D.F. given in equation (V.2), the classification of an individual may also be made to depend on the true value of π_1. This is acheived by allowing the constant that the classification score, S, must be greater than to declare an individual as LESA to depend on π_1; see subsection F.1.)

To illustrate the C procedure for the children field test data, RTI considered the 12 cell table defined by recoding Spund (9 categories),

Usage (4 categories) and Years (7 categories) as follows:

Spund' = 1 if Spund = 2 thru 7
2 if Spund = 8, 9 or 10

Usage' = 1 if Usage = 0, 1, 2
2 if Usage = 3

Years' = 1 if Years = 0, 1
2 if Years = 2
3 if Years ≥ 3

For this 12 cell table, RTI estimated the probabilities of LESA and Non-LESA from a crosstab of LSTCD by Spund by Usage by Years. Then using these cell probabilities the C procedure's classification rule based upon equation (V.3) was applied for two different estimates of π_1, π_2 (i) π_1 and π_2 as observed in the field test data and (ii) $\pi_2 = 3\pi_1$. The results of these computations are given in Table V.20. The table also shows that the results of using the π_1, π_2 as observed classification rule on the sample field test data are that 75% of the children are correctly classified. This compares with 76.1% when using the 3 variable classical D.F. given in Section F.1 (see Table V.17). Note that the C procedure's rule for classifying the field test data when π_1 and π_2 are as observed is nothing more than classifying a cell as LESA when there are more LESAs than Non-LESAs in the cell as defined by LSTCD. (Also note for this case that if the number of LESAs = number of Non-LESAs in a cell then a coin toss may be used to break the tie.)

To further investigate the effect of using the D.F. versus the C procedure, RTI performed the same analysis as given in Table V.20 using FCTR to define LESA. In addition, RTI also used the C procedure for the case when Spund, Usage and Years had 9, 4 and 7 categories respectively

Table V.20 Illustration of Cochran's Classification Procedure for the Children Field Test Data for Two Different Estimates of Π_1, Π_2

Value of Discriminating Variables			LSTCD Used to Define LESA		Classification Rule When Π_1, Π_2 as observed[1]	Classification Rule if $\Pi_2 = 3\,\Pi_1$
Usage'	Spund'	Years'	Cell Prob. for LESA $= P_i$	Cell Prob. for Non-LESA $= P'_i$		
1	1	1	.316	.039	LESA	LESA
1	1	2	.120	.024	LESA	LESA
1	1	3	.111	.031	LESA	LESA
1	2	1	.123	.101	LESA	Non-LESA
1	2	2	.092	.070	LESA	Non-LESA
1	2	3	.178	.274	Non-LESA	Non-LESA
2	1	1	.003	.004	LESA	Non-LESA
2	1	2	.002	.002	LESA	Non-LESA
2	1	3	.002	.000	LESA	LESA
2	2	1	.009	.162	Non-LESA	Non-LESA
2	2	2	.011	.136	Non-LESA	Non-LESA
2	2	3	.034	.151	Non-LESA	Non-LESA
		Total	1.00	1.00	[2]	

[1] Observed $\Pi_1 = .585$, $\Pi_2 = .415$ (see Table V.16)

[2] When the sample field test data was classified using the π_1, π_2 as observed rule, the results were 823/1098 = .75 of the children were correctly classified.

(i.e., the number of levels for the 3 variables were not reduced by recoding as in Table V.20). This resulted in applying the C procedure to a 252 cell table (9x4x7 = 252). The prediction results of these computations using the field test data and assuming $\pi_1 = .585$ are given in Table V.21. Note that the C procedure applied to the 252 cell table essentially indicates the best we can expect the variables Spund, Usage and Years to do in correctly classifying children in the field test as LESA and Non-LESA (as defined by LSTCD or FCTR).

Table V.21 indicates that the D.F. does about as well in classifying children as the two discrete classification rules. In addition, the 12 cell rule does almost as well as the 252 cell rule. Thus, for the data examined here, both the 12 cell rule and the D.F. appear to efficiently use the classification information on LESA contained in Spund, Usage and Years.

F.4 Estimation of Percent LESA

Because of the fact that the MELP based on the Census Questions will be used to estimate the percent LESA in the general population, RTI also examined how the estimate based on the D.F. might be obtained in practice. (Recall that estimates of % LESA based on the proportion of individuals classified by the D.F. as LESA have been given in the tables presented in subsections F.1, F.2 and F.3). Before presenting numerical results the following general comments should be considered:

In general, for the two population classification problem there are two errors of misclassification, α_1 and α_2, which have been defined in Table V.16. In classical discriminant analysis the discriminant function (equation (V.2)) is estimated by minimizing $\alpha_1 + \alpha_2$. Now if one

Table V.21 Comparison of Prediction Results Over Sites Using Different Classification Procedures to Predict Whether a Child is LESA or Non-LESA; Procedures Based on Spund, Usage and Years; LESA Defined by LSTCD and FCTR[1/]

LSTCD Used to Define LESA Over Sites

LSTCD Group	252 Cell Procedure[2/4/] Predicted Group LESA	Non-LESA	12 Cell Procedure Predicted Group LESA	Non-LESA	Discriminant Analysis Predicted Group LESA	Non-LESA
LESA	528	114	490	152	488	154
Non-LESA	97	359	123	333	108	348
α_1	21.6		23.6		24.0	
α_2	21.3		27.0		23.7	
% Correct	80.8		75.0		76.1	
Est % LESA	56.9		55.8		54.3	
Actual % LESA	58.5		58.5		58.5	

FCTR[3/] Used to Define LESA Over Sites

FCTR Group	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA
LESA	533	107	502	138	471	169
Non-LESA	95	363	112	346	99	359
α_1	16.7		21.6		26.4	
α_2	20.7		24.5		21.6	
% Correct	81.6		77.2		75.6	
Est % LESA	57.2		55.9		51.9	
Actual % LESA	58.3		58.3		58.3	

1/ Definitions of α_1, α_2, % correct, Est % LESA and actual % LESA given in Table V.16.

2/ 252 cells = Spund (9 levels) x Usage (4 levels) x Years (7 levels)
12 cells = Spund' (2 levels) x Usage' (2 levels) x Years' (3 levels).
The 252 cell and 12 cell procedures classify the i^{th} cell as LESA if $\Pi_1 P_i > \Pi_2 P'_i$.

3/ Over sites FCTR ≤.43 = LESA.

4/ π_1 and π_2 assumed to be .585 and .415, respectively, when applying the 252 and 12 cell procedures.

defines, as in subsection F.3,

π_1 = frequency of LESAs in population,

π_2 = frequency of Non-LESAs in population,

then we have the following table after predicting LESA and Non-LESA by the D.F.

Table V.22 Schematic of Misclassification Errors

Actual Group	Predicted Group by D.F. LESA	Non-LESA	Total
LESA	$(1-\alpha_1)\pi_1$	$\alpha_1\pi_1$	π_1
Non-LESA	$\alpha_2\pi_2$	$(1-\alpha_2)\pi_2$	π_2
Total	$(1-\alpha_1)\pi_1 + \alpha_2\pi_2$	$\alpha_1\pi_1 + (1-\alpha_2)\pi_2$	1.00

The usual estimate of the frequency of LESA from Table V.22, as given in previous subsections, would be $(1-\alpha_1)\pi_1 + \alpha_2\pi_2$. Now consider this estimate for the following three cases based on the relative sizes of π_1, π_2 and α_1, α_2:

(i) if $\pi_1 = \pi_2$ (i.e., frequency of LESA = frequency of Non-LESA) and $\alpha_1 = \alpha_2$ (i.e., errors of misclassification are equal) then the estimate of the frequency of LESA from Table V.22 is:

$$(1-\alpha_1)\pi_1 + \alpha_2\pi_2 = (1-\alpha_1)\pi_1 + \alpha_1\pi_1 = \pi_1$$

(ii) if $\pi_2 = k\pi_1$ and $\alpha_1 = \alpha_2$ then the estimate of the frequency of LESA is:

$$(1-\alpha_1)\pi_1 + \alpha_2\pi_2 = \pi_1 + \alpha_1(k-1)\pi_1$$

(Here, if α_1 and k are relatively large, then the bias in the estimate of π_1 can be quite high.)

(iii) if $\pi_2 = k\pi_1$ and $\alpha_2 = \frac{\alpha_1}{k}$ then the estimate of the frequency of LESA is:

$$(1-\alpha_1)\pi_1 + \alpha_2\pi_2 = (1-\alpha_1)\pi_1 + \frac{\alpha_1}{k}k\pi_1 = \pi_1.$$

Thus, it is obvious that the accuracy of our estimate of the frequency of LESA not only depends on the errors of misclassification but also on the relative magnitudes of π_1 and π_2 in the population.

The present sample field test data for children over sites has a value of π_1 equal to .585 when LESA is defined by LSTCD. However, according to CAL and National Center for Educational Statistics staff, the actual value of π_1 should be closer to .2 for the sample population that the MELP will be applied to. Accordingly, RTI adjusted the sample field test data for children in the following way to determine the effect on the usual estimate of % LESA when the number of Non-LESAs was much larger than the number of LESAs:

(i) If a child had LSTCD = 0 (LESA) then the results for that child were written onto a new data file only once,

(ii) If a child had LSTCD = 1 (Non-LESA) then the results for that child were written onto the new data file four times.

The results of this adjustment were a new data file with $\pi_1 = .26$ and $\pi_2 = .74$ over sites. RTI then performed discriminate analyses on this new data file and the results are given in Tables V.23, V.24 and V.25. Table V.23 presents the results when LSTCD is used to define LESA. The table shows that in all cases (by sites and over sites) the estimate of % LESA is larger than the actual % LESA as defined by LSTCD. This is in contrast to the results before adjustment (i.e., $\pi_1 = .585$ and $\pi_2 = .415$) when the estimate of % LESA is usually smaller than the

(Text continued on page 93)

Table V.23 Results of Using Discriminant Analysis to Predict Whether A Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, Field Data Adjusted so that Over Sites % LESA =26%, % Non-LESA = 74% as Defined by LSTCD[1/] [3/]

LSTCD Used to Define LESA

LSTCD Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	146	64	156	35	33	20	66	29	64	29
Non-LESA	80	348	88	604	96	224	60	112	60	152
α_1 [2/]	30.5 (23.8)		18.3 (15.7)		37.7 (39.6)		30.5 (31.6)		31.2 (25.8)	
α_2	18.7 (25.2)		12.7 (15.6)		30.0 (28.8)		34.9 (37.2)		28.3 (30.2)	
% Correct	77.4 (75.7)		86.1 (84.3)		68.9 (66.9)		66.7 (66.7)		70.8 (72.6)	
Est LESA	35.4 (59.0)		27.6 (51.6)		34.6 (41.4)		47.2 (58.7)		40.7 (58.2)	
Actual LESA	32.9 (66.2)		21.6 (52.5)		14.2 (39.8)		35.6 (68.8)		30.5 (63.7)	

Over Sites

LSTCD Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	465	177
Non-LESA	356	1468
α_1	27.6 (24.0)	
α_2	19.5 (23.7)	
% Correct	78.4 (76.1)	
Est LESA	33.3 (54.3)	
Actual LESA	26.0 (58.5)	

[1/] Discriminant function computed for each site.

[2/] Definitions of α_1, α_2, % correct, Est LESA and actual LESA given in Table V.16.

[3/] Numbers in ( ) give results when unadjusted field test data was used.

Table V.24 Standardized and Unstandardized Discriminant Function Coefficients for Children for Spund, Usage and Years, Field Test Data Adjusted so that Over Sites % LESA = 26%, % Non-LESA = 74% as Defined by LSTCD

	LSTCD Used to Define LESA					
Discriminating Variables	Site 1 (Miami)	Site 3 (El Paso)	Site 4 (S.F. Non-Chinese)	Site 5 (Ganado)	Site 6 (S.F. Chinese)	Over Sites
	Standardized Coefficients 1/ 2/					
Spund	-.62 (-.49)	-.59 (-.43)	-.37 (-.36)	-.44 (-.39)	-.54 (-.46)	-.59 (-.47)
Usage	-.13 (-.12)	-.49 (-.64)	-.76 (-.76)	-.68 (-.74)	-.51 (-.54)	-.51 (-.62)
Years	-.46 (-.59)	-.03 (.01)	-.24 (-.16)	-.16 (-.08)	-.17 (-.24)	-.16 (-.10)
	Unstandardized Coefficients					
Spund	-.30 (-.22)	-.29 (-.18)	-.25 (-.21)	-.20 (-.17)	-.26 (-.21)	-.29 (-.21)
Usage	-.17 (-.15)	-.41 (-.50)	-.91 (-.82)	-.55 (-.64)	-.49 (-.53)	-.44 (-.54)
Years	-.25 (-.34)	-.04 (.01)	-.15 (-.10)	-.08 (-.04)	-.12 (-.18)	-.09 (-.06)
Constant	3.31 (2.53)	3.25 (2.04)	4.82 (3.88)	2.62 (2.09)	3.14 (2.62)	3.33 (2.36)

1/ Standardized coefficients indicate the relative contribution of each variable to the discriminant function.
Unstandardized coefficients are multiplied by the raw values of Spund, Usage and Years to obtain a discriminant score which is used to classify a child as LESA or Non-LESA.

2/ Numbers in ( ) give results when unadjusted field test data was used.

Table V.25 Results of Using Discriminant Analysis to Predict Whether a Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, Field Test Data Adjusted so that Over Sites % LESA = 39%, % Non-LESA = 61% as Defined by FCTR[1/]

FCTR Used to Define LESA[3/4/]

Over Sites			Over Sites
FCTR Group	Predicted Group LESA	Non-LESA	Discriminating Variables	Standardized Coefficients
LESA	614	357	Spund	-.50 (-.44)
Non-LESA	372	1123	Usage	-.43 (-.47)
			Years	-.47 (-.42)
α_1 [2/]	36.8 (26.4)			
α_2	24.9 (21.6)			Unstandardized Coefficients
% Correct	70.4 (75.6)			
Est LESA	40.0 (51.9)		Spund	-.25 (-.19)
Actual LESA	39.4 (58.3)		Usage	-.37 (-.41)
			Years	-.28 (-.26)
			Constant	3.35 (2.56)

[1/] Table gives prediction results and discriminant function coefficients for FCTR.

[2/] Definitions of α_1, α_2, % correct, Est LESA and actual LESA given in Table V.16

[3/] Numbers in ( ) give results when unadjusted field test data was used.

[4/] Over sites FCTR ≤.43 = LESA.

actual % LESA. Thus, in the present case the effect of reducing the proportion of LESAs in the sample population has resulted in an overestimate of % LESA by the D.F. Examination of Table V.23 indicates that the cause of the overestimation is the number of Non-LESAs predicted to be LESAs (i.e., $\alpha_2 \pi_2$ in Table V.22). That is, it appears that even though α_2 has been reduced somewhat from the unadjusted case (e.g., over sites α_2 was reduced from 23.7% to 19.5%) the large increase in π_2 has caused an overestimate of the % LESA. Table V.25 indicates essentially the same conclusion when FCTR is used to define LESA. Table V.30 in subsection F.5 also given similar results using the adult field test data.

Because of the problem in estimating the % LESA as the proportion of predicted LESAs, RTI considered two other possible estimates of π_1. The first estimate can be obtained for a future sample by equating the observed proportion of LESA in the future sample (call it L where 100xL is the estimate of % LESA that has been used previously in this section) to the expected frequency of LESA in the sample from Table V.22; i.e.,

$$L = (1-\alpha_1)\pi_1 + \alpha_2(1-\pi_1)$$

where α_1 and α_2 are estimated from the sample field test data. Solving for π_1 gives

$$\hat{\pi}_1 = \frac{L-\alpha_2}{1-\alpha_1-\alpha_2} \qquad (V.4)$$

Note that $\hat{\pi}_1$ is an unbiased estimate of π_1 if α_1 and α_2 are known. Also note, if $\hat{\alpha}_1 = \hat{\alpha}_2 = .25$ and L is less than .5 as is the case for the adjusted field test data, that equation (V.4) gives $\hat{\pi}_1 < L$.

Using equation (V.4), RTI estimated the % LESA for the cases given in Table V.23 by using the L observed in the table and α_1 and α_2 from the unadjusted field test data (given in ( ) in the table). For example, over

sites the computation is

$$\hat{\pi}_1 = \frac{.333-.237}{1-.24-.237} = .184$$

The results are as follows:

	Site 1	Site 3	Site 4	Site 5	Site 6	Over Sites
Est % LESA by Eq. (V.4)	20.0	17.5	18.4	32.1	23.9	18.4
Est % LESA Table V.23	35.4	27.6	34.6	47.2	40.7	33.3
Actual % LESA	32.9	21.6	14.2	35.6	30.5	26.0

In general, compared to the usual estimate of % LESA, equation (V.4) gives a more accurate estimate in Sites 3, 4, 5 and 6 and an equally accurate estimate over sites. Thus, equation (V.4) may be a better method for estimating the % LESA than the usual method of just taking the predicted number of LESAs, 100xL. However, two major drawbacks of equation (V.4) should be noted here,

(i) if $\alpha_2 > L$ then $\hat{\pi}_1$ will be negative, (This is a real problem in the present case because of the relatively large magnitudes of α_1 and α_2)

(ii) Poor estimates of α_1 and α_2 from the sample field test data can result in a very inaccurate estimate of π_1. (Recall that the estimates of α_1 and α_2 given throughout this section are biased estimates (i.e., the estimates are too optimistic); and in addition, note from Table V.23 how these estimates varied for the unadjusted data versus the adjusted data.)

Accordingly, RTI would suggest that a better estimate of the % LESA than equation (V.4) can be obtained by utilizing a double sampling procedure in the following manner:

(i) Classify the individuals in the Survey of Income and Education administered by the Census Bureau as LESA or Non-LESA by use of a D.F. based on the CQs. Denote this sample size by N.

(ii) Subsample the Census sample and in this subsample determine whether each individual is LESA or Non-LESA by either obtaining school lists (i.e., so that LSTCD may be determined) or by administrating the criterion packages (i.e., so that FCTR may be determined). Denote this sample size by n. Thus, for the subsample, individuals are classified by LSTCD or FCTR and the D.F. while for the rest of the sample of size N-n, individuals are only classified by the D.F.

The resulting data are as follows:

FCTR or LSTCD Group	Predicted Group by D.F.		Sample Size
	LESA	Non-LESA	
LESA	n_{00}	n_{01}	
Non-LESA	n_{10}	n_{11}	
	$n_{.0}$	$n_{.1}$	n
	100 x L	100 x NL	N-n
Totals	(100xL) + $n_{.0}$	(100xNL) + $n_{.1}$	N

where 100 x L is the number of individuals classified as LESA by the D.F. in the SIE sample minus the subsample (i.e., sample size N-n) and the n_{ij}'s are the sample sizes observed in the subsample of size n.

(iii) Obtain an estimate of the frequency of LESA in the population by the following formula:

$$\hat{\pi}_1 = \frac{n_{00}}{n_{.0}} \frac{(100xL) + n_{.0}}{N} + \frac{n_{01}}{n_{.1}} \frac{(100xNL) + n_{.1}}{N} \qquad (V.5)$$

The above estimate which is given by Tenenbein [6] is the maximum likelihood estimate of π_1. The estimate essentially takes the proportion of the N individuals which have been classified as LESA or Non-LESA by the D.F. (i.e., $(100xL + n_{.0})/N$ and $(100xNL + n_{.1})/N$) and corrects by multiplication by the ratios $n_{00}/n_{.0}$ and $n_{01}/n_{.1}$, respectively.

RTI feels that the above procedure should produce a reliable estimate of the frequency of LESA in the general population.

F.5 Discriminant Functions for Adults

In addition to the discriminant analyses on the field test data for children given previously, RTI also carried out discriminant analyses on the adult field test data. These adult analyses are quite similar to the child analyses; and therefore, references to the corresponding child computations are given throughout this subsection. For the adult analyses LESA and Non-LESA groups were defined by LSTCD and FCTR. Recall that there were no lists for adults in Sites 4, 5 and 6; and therefore, in these sites LSTCD was obtained by classifying an adult as LSTCD = 0 (i.e. LESA) if their child had LSTCD = 0 or if the adult had FCTR $\leq$.1. In all sites, an adult was classified by FCTR as LESA if FCTR $\leq$.1. The constant, .1, was determined so that the number of LESAs defined by FCTR in Site 1 (Miami) was approximately equal to the number of LESAs defined by LSTCD in Site 1. In this subsection the following analyses of the adult data are described: discriminate analyses by site and over sites using both 11 and 6 CQs as discriminators; comparison of the over site D.F. versus the within site D.F.s; discrimination using Cochran's discrete rule; and discriminate analyses on adjusted adult data where the proportion of LESAs is adjusted to equal .19.

The results of using discriminant analyses to classify the adult field test data using 11 CQs as discriminators is given in Table V.26. The format of the table is the same as that used in previous subsections for children with the estimated % LESA again being the proportion of adults classified as LESA by the D.F.s. The 11 CQs are the same questions used in Section E as the independent regression variables (see Table V.14). Table V.26 indicates that over sites 83% of the adults are classified the

same by the D.F. and FCTR while the corresponding value for LSTCD is 64%. In general, the agreement in classification of LESA and Non-LESAs by the D.F.s and FCTR is better than for the D.F.s and LSTCD (this same result was also evident for the multiple regression analyses described in Section E). In addition, it appears for the adults data that the agreement in classification results between the D.F.s and FCTR is somewhat higher than was the case for the child field test data (see Table V.16).

After examining the results of the 11 CQs discriminate analyses and multiple regression analyses (given in Section E), RTI also computed D.F.s for adults based only on the five variables Spund = Speak + Und, Years, News, Birth and Grade. The results of these computations are given in Table V.27. The numbers in ( ) in the table are the corresponding results when 11 CQs were used to obtain D.F.s. The table indicates that the D.F.s based on the 5 variables do about as well in classifying adults as the D.F.s based on 11 CQs. This is particularly true when FCTR is used to define LESA/Non-LESA groups. Thus, it would appear that a subset of the 11 CQs will classify adults as LESA/Non-LESA almost as well as the full set of 11 Census Questions.

Recall in subsection F.2 that a comparison was made between the classification of children by the D.F. computed over sites and the D.F.s computed within each site (see Table V.19). The reason for this comparison was to investigate the possibility of using only one D.F. over all sites. Accordingly, RTI also made this comparison for the adult D.F.s and the results are given in Table V.28. The table only contains results for LESA defined by FCTR. Table V.28 indicates that the over site D.F. certainly does as well overall in correctly classifying adults as do the

within site D.F.s. However, as was the case for children (see Table V.19) the estimated magnitude of the errors of misclassification are quite different in most sites for the two types of D.F.s. For example, in Site 1, α_1 = 9.7% and α_2 = 60.9% for the over site D.F. while the corresponding percentages for the within site D.F. are 23.2% and 26.4%. In fact, as one might expect, it is always the case that the estimated values of α_1 and α_2 are closer in magnitude for the within site D.F. versus the over site D.F. Recall that the implications of differences in α_1 and α_2 on the estimate of % LESA have been discussed in subsection F.4. Table V.28 also shows that the estimated % LESAs in each site are quite different for the over site D.F. versus the within site D.F.s.

Since the discriminating variables (i.e., the CQs) are discrete in the present case, RTI used Cochran's discrete discriminating rule to classify the adult field test data as was done in subsection F.3 for the child field test data. To do this a 4 cell table was defined by recoding Spund (9 categories) and Years (7 categories) into 2 categories each; namely, Spund <8 and ≥ 8; and Years ≤ 3 and >3. For this 4 cell table, RTI estimated the probabilities over sites of LESA and Non-LESA from the adult field test data. Cochran's procedure based upon equation (V.3) in subsection F.3 was then applied assuming that π_1 and π_2 were as observed in the field test. The results of these calculations are given in Table V.29. The table indicates that when Cochran's procedure is used over sites, 83% of the adults are classified the same as by FCTR. Recall from Table V.26 that the corresponding percentage was 83% when a classical D.F. based on 11 CQs was utilized. Thus, over sites, Cochran's procedure using only two categories of both Spund and Years agrees with the classification of LESA/Non-LESA groups by FCTR as well as the classical D.F. based on 11 CQs.

Finally, as in subsection F.4 for children, RTI adjusted the sample field test data for adults so that the number of Non-LESAs was much larger than the number of LESAs as defined by FCTR. Recall that was done to determine how discriminant analysis would perform for the case when the proportion of LESAs was close to .2. In particular, the adjustment for adults was done in the following manner:

(i) If an adult had FCTR ≤.1 (i.e., LESA) then the results for that adult were written onto a new data file only once,

(ii) If an adult had FCTR >.1 (i.e., Non-LESA) then the results for that adult were written onto the new data file four times.

The results of this adjustment were a new data file with $\pi_1 = .19$ and $\pi_2 = .81$ over sites (as defined by FCTR). Discriminate analyses were then performed on this new data file and the results are given in Table V.30. The table indicates results which are similar to those for the adjusted children field test data (see Table V.23). That is, the estimate of % LESA (as given by the proportion of adults predicted to be LESA by the D.F.s) is larger in all cases (by sites and over sites) than the actual % LESA as defined by FCTR. Again, this overestimation is due to the relatively large number of Non-LESAs predicted to be LESAs.

(Text continued on page 108)

Table V.26 Results of Using Discriminant Analysis to Predict Whether an Adult is LESA or Non-LESA, Discriminant Function Based on 11 Census Questions, LESA Group Defined by LSTCD and FCTR[1/2/]

LSTCD used to define LESA[3/]

	Site 1 (Miami)		Site 3 (El Paso)		Site 4 (S.F. Non-Chinese)		Site 5 (Ganado)		Site 6 (S.F. Chinese)	
LSTCD Group	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA
LESA	135	49	54	38	41	15	54	19	77	4
Non-LESA	22	66	47	63	19	41	9	20	10	20
α_1 [4/]	26.6		41.3		26.8		26.0		4.9	
α_2	25.0		42.7		31.7		31.0		33.3	
% Correct	73.9		57.9		70.7		72.6		87.4	
Est % LESA	57.7		50.0		51.7		61.8		78.4	
Actual % LESA	67.6		45.5		48.3		71.6		73.0	

Over Sites

LSTCD Group	Predicted Group LESA	Non-LESA
LESA	330	156
Non-LESA	131	186
α_1	32.1	
α_2	41.3	
% Correct	64.3	
Est % LESA	57.4	
Actual % LESA	60.5	

Table V.26 Continued

FCTR used to define LESA[5/]

FCTR Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (N.E. Arizona) Predicted Group LESA	Site 5 (N.E. Arizona) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	142	43	135	25	12	5	20	6	49	7
Non-LESA	23	64	10	32	20	79	29	159	16	39
α_1	23.2		15.6		29.4		23.1		12.5	
α_2	26.4		23.8		20.2		15.4		29.1	
% Correct	75.7		82.7		78.5		83.6		79.3	
Est % LESA	60.7		71.8		27.6		22.9		58.6	
Actual % LESA	68.0		79.2		14.7		23.4		50.5	

Over Sites

Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	389	55
Non-LESA	102	369
α_1	12.4	
α_2	21.7	
% Correct	82.8	
Est % LESA	53.7	
Actual % LESA	48.5	

1/ 11 Census Questions = When, Speak, Und, Kid, Frnd, Hlang, Years, News, Birth, Grade and Incm.

2/ Discriminant Function computed for each site.

3/ In Site 4, 5 and 6 LSTCD for adults was obtained by classifying an adult as LESA (LSTCD = 0) if their child was LESA or if the adult had a FCTR score ≤.1.

4/ α_1 = Percent of LESAs classified as Non-LESA; α_2 = Percent of Non-LESAs classified as LESA; % correct = overall percent classified correctly.
Est % LESA = estimated percent LESA = 100 x (Number predicted as LESA)/(Total number of children)
Actual % LESA = percent LESAs as given by LSTCD (or FCTR).

5/ FCTR ≤.1 = LESA in all sites.

Table V.27 Results of Using Discriminant Analysis to Predict Whether an Adult is LESA or Non-LESA, Discriminant Function Based on 6 Census Questions, LESA Group Defined by LSTCD and FCTR [1/] [2/] [6/]

LSTCD used to define LESA[3/]

LSTCD Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (Ganado) Predicted Group LESA	Site 5 (Ganado) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	129	55	43	49	37	19	47	26	69	12
Non-LESA	21	67	37	73	16	44	9	20	8	22
α_1 [4/]	29.9 (26.6)		53.3 (41.3)		33.9 (26.8)		35.6 (26.0)		14.8 (4.9)	
α_2	23.9 (25.0)		33.6 (42.7)		26.7 (31.7)		31.0 (31.0)		26.7 (33.3)	
% Correct	72.1 (73.9)		57.4 (57.9)		69.8 (70.7)		65.7 (72.6)		82.0 (87.4)	
Est % LESA	55.1 (57.7)		39.6 (50.0)		45.7 (51.7)		54.9 (61.8)		69.4 (78.4)	
Actual % LESA	67.6		45.5		48.3		71.6		73.0	

Over Sites

LSTCD Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	291	195
Non-LESA	118	199
α_1	40.1 (32.1)	
α_2	37.2 (41.3)	
% Correct	61.0 (64.3)	
Est % LESA	50.9 (57.4)	
Actual % LESA	60.5	

Table V.27 Continued

FCTR used to define LESA[5/]

FCTR Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (N.E. Arizona) Predicted Group LESA	Site 5 (N.E. Arizona) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	138	47	134	26	11	6	19	7	50	6
Non-LESA	25	62	14	28	20	79	25	163	17	38
α_1	25.4 (23.2)		16.3 (15.6)		35.3 (29.4)		26.9 (23.1)		10.7 (12.5)	
α_2	28.7 (26.4)		33.3 (23.8)		20.2 (20.2)		13.3 (15.4)		30.9 (29.1)	
% Correct	73.5 (75.7)		80.2 (82.7)		77.6 (78.5)		85.1 (83.6)		79.3 (79.3)	
Est % LESA	59.9 (60.7)		73.3 (71.8)		26.7 (27.6)		20.6 (22.9)		60.4 (58.6)	
Actual % LESA	68.0		79.2		14.7		23.4		50.5	

Over Sites

FCTR Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	385	59
Non-LESA	105	366
α_1	13.3 (12.4)	
α_2	22.3 (21.7)	
% Correct	82.1 (82.8)	
Est % LESA	53.6 (53.7)	
Actual % LESA	48.5	

1/ 6 Census Questions = Spund = Speak + Und; Years; News; Birth; Grade

2/ Discriminant Function computed for each site.

3/ In Site 4, 5 and 6 LSTCD for adults was obtained by classifying an adult as LESA (LSTCD = 0) if their child was LESA or if the adult had a FCTR score ≤.1.

4/ α_1 = Percent of LESAs classified as Non-LESA; α_2 = Percent of Non-LESAs classified as LESA; % correct = overall percent classified correctly.
Est % LESA = estimated percent LESA = 100 x (Number predicted as LESA)/(Total number of children)
Actual % LESA = percent LESAs as given by LSTCD (or FCTR).

5/ FCTR ≤.1 = LESA in all sites.

6/ Numbers in ( ) give results when 11 CQs were used to compute the discriminate function (see Table V.26)

Table V.28 Results of Using a Discriminant Function Estimated Over All Sites to Predict Whether an Adult is LESA or Non-LESA Within Each Site, Discriminant Function Based on 11 Census Questions, LESA Defined by FCTR[1/] [3/]

	FCTR[2/] Used to Define LESA									
	Site 1 (Miami)		Site 3 (El Paso)		Site 4 (S.F. Non-Chinese)		Site 5 (N.E. Arizona)		Site 6 (S.F. Chinese)	
FCTR Group	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA
LESA	167	18	155	5	9	8	19	7	51	5
Non-LESA	53	34	25	17	5	94	15	173	20	35
α_1	9.7 (23.2)		3.1 (15.6)		47.1 (29.4)		26.9 (23.1)		8.9 (12.5)	
α_2	60.9 (26.4)		59.5 (23.8)		5.1 (20.2)		8.0 (15.4)		36.4 (29.1)	
% Correct	73.9 (75.7)		85.1 (82.7)		88.8 (78.4)		89.7 (83.6)		77.5 (79.3)	
Est % LESA	80.9 (60.7)		89.1 (71.8)		12.1 (27.6)		15.9 (22.9)		64.0 (58.6)	
Actual % LESA	68.0		79.2		14.7		12.1		50.5	

105

1/ Definitions of α_1, α_2, % correct, Est % LESA and actual % LESA given in Table V.16.

2/ FCTR ≤.1 = LESA in all sites.

3/ Numbers in ( ) give results when discriminant functions were computed for each site (see Table V.26)

Table V.29. Cochran's Classification Procedure for the Adult Field Test Data Over Sites

Value of Discriminating Variables		FCTR[1/] Used to Define LESA		Classification Rule When π_1, π_2 as Observed[2/]
		Cell Prob. for LESA	Cell Prob. for Non-LESA	
SPUND	YEARS	$= P_i$	$= P_i'$	
<8	≤3	.818	.166	LESA
<8	>3	.052	.106	Non-LESA
≥8	≤3	.086	.149	Non-LESA
≥8	>3	.045	.580	Non-LESA
TOTAL		1.00	1.00	[3/]

[1/] FCTR ≤.1 = LESA.

[2/] Observed π_1 = .485; π_2 = .515 (see Table V.26).

[3/] When the sample field test data was classified using the π_1, π_2 as observed rule, the results were 756/915 = .83 of the adults were correctly classified (where correct means classified the same as FCTR).

Table V.30 Results of Using Discriminant Analysis to Predict Whether an Adult is LESA or Non-LESA, Discriminant Function Based on 11 Census Questions, Field Data Adjusted so that Over Sites % LESA = 19%, % Non-LESA = 81% as Defined by FCTR [1/] [3/]

FCTR Used to Define LESA [4/]

	Site 1 (Miami)		Site 3 (El Paso)		Site 4 (S.F. Non-Chinese)		Site 5 (N.E. Arizona)		Site 6 (S.F. Chinese)	
FCTR Group	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA
LESA	138	47	132	28	11	6	20	6	47	9
Non-LESA	88	260	40	128	68	328	116	636	52	168
α_1 [2/]	25.4 (23.2)		17.5 (15.6)		35.3 (29.4)		23.1 (23.1)		16.1 (12.5)	
α_2	25.3 (26.4)		23.8 (23.8)		17.2 (20.2)		15.4 (15.4)		23.6 (29.1)	
% Correct	74.7 (75.7)		79.3 (82.7)		82.1 (78.5)		84.3 (83.6)		77.9 (79.3)	
Est % LESA	42.4 (60.7)		52.4 (71.8)		19.1 (27.6)		17.5 (22.9)		35.9 (58.6)	
Actual % LESA	34.7 (68.0)		48.8 (79.2)		4.1 (14.7)		3.3 (23.4)		20.3 (50.5)	

Over Sites

FCTR Group	Predicted Group LESA	Non-LESA
LESA	365	79
Non-LESA	368	1516
α_1	17.8 (12.4)	
α_2	19.5 (21.7)	
% Correct	80.8 (82.8)	
Est % LESA	31.5 (53.7)	
Actual % LESA	19.1 (48.5)	

[1/] Discriminant function computed for each site.

[2/] Definitions of α_1, α_2, % correct, Est % LESA and actual % LESA given in Table V. 16.

[3/] Number in ( ) give results when unadjusted field test data was used.

[4/] FCTR ≤.1 = LESA in all sites.

G. Reverse Scale Questions

In San Francisco and Arizona during the last week of the field work RTI asked four of the questions on the Census Questionnaire in two different ways. In particular, the scales for the questions asking how well does the individual speak or understand English or another language (i.e. question numbers 6,7,10 and 11, see Appendix A) were reversed for some of the sample respondents to determine if this effected the distribution of responses. Thus, during the last week some respondents were administered the usual pink Census Questionnaire which had the four questions with the first response as "very well" while the remaining respondents were administered a special form (see Exhibit V.1) where the response choices for the four questions were reversed ("not at all" was first).

The results of comparing the two scales (normal and reversed) for the four questions in the two sites for adults and children are given in Table V.31. The table presents the distribution of respondents for the normal and reversed scales for each question. In addition, the results of a χ^2 test are presented which tests the hypothesis that the distribution of respondents is the same for the normal and reversed scales. The table shows that out of the 16 χ^2 tests performed only one test is significant (at the .01 level). Thus, the data in Table V.31 do not indicate that the distribution of respondents for the two scales for the four Census Questions are different. However, it should be pointed out here that the number of respondents (particularily for the reversed scale) were quite limited in most cases making the power (i.e. ability to detect differences in the distibutions) of the χ^2 test quite small.

(Text continued on page 113)

BILINGUAL STUDY

CENSUS QUESTIONNAIRE

O.M.B. 51-S75048
Expires October 1975

(Items 6, 7, 10, 11 have reversed scales)

ID No. of DR _____ Sex _____

FI _____ FI No. _____ Date ____

Type (✓): ☐ Self Report ☐ Second Hand Report

1. What is . . .'s date of birth?

   Month _____ Day _____ Year _____

2. What is . . .'s origin or descent? *(USE FLASH CARD A)*

   _____

3. In what state, U. S. territory, or foreign country was . . . born? *(USE FLASH CARD B)*

   _____

   (IF ANSWER IS "this state" OR "different state" SKIP TO Q.5.)

4. When did . . . come to the U. S. mainland to stay?

   _____ 1. 1975
   _____ 2. 1973-1974
   _____ 3. 1971-1972
   _____ 4. 1966-1970
   _____ 5. 1961-1965
   _____ 6. Before 1961
   _____ 7. Don't know

5. Does . . . speak or understand <u>any</u> English?

   _____ 1. Yes
   _____ 2. No *(SKIP TO Q.8)*
   _____ 3. Don't know *(SKIP TO Q.8)*

6. How well does . . . speak English? *(READ ANSWER CHOICES 1-5)*

   _____ 1. Not at all
   _____ 2. Just a little
   _____ 3a. Adequately only for a few purposes
   _____ 3b. Adequately for most purposes
   _____ 4. Well
   _____ 5. Very well
   _____ 6. Don't know

7. How well does . . . understand spoken English? *(READ ANSWER CHOICES 1-5)*

   _____ 1. Not at all
   _____ 2. Just a little
   _____ 3a. Adequately only for a few purposes
   _____ 3b. Adequately for most purposes
   _____ 4. Well
   _____ 5. Very well
   _____ 6. Don't know

8. What *(OTHER)* languages does . . . speak? *(USE FLASH CARD C)*

   _____

   (IF NONE, SKIP TO Q.12. IF ONLY ONE, SKIP TO Q.10)

9. Which one of these languages does . . . speak most often? *(USE FLASH CARD C)*

_____

10. How well does . . . speak *(PRINCIPAL LANGUAGE FROM Q.8 OR Q.9)*? *(READ ANSWER CHOICES 1-4)*

_____ 1. Just a little
_____ 2a. Adequately only for a few purposes
_____ 2b. Adequately for most purposes
_____ 3. Well
_____ 4. Very well
_____ 5. Don't know

11. How well does . . . understand *(PRINCIPAL LANGUAGE FROM Q.8 OR Q.9)*? *(READ ANSWER CHOICES 1-4)*

_____ 1. Just a little
_____ 2a. Adequately only for a few purposes
_____ 2b. Adequately for most purposes
_____ 3. Well
_____ 4. Very well
_____ 5. Don't know

12. What language does . . . usually speak when talking to: *(USE FLASH CARD C)*

a. brothers and sisters? _____
b. parents? _____
c. other older relatives? _____
d. . . .'s best friend? _____
e. *(IF . . . IS AN ADULT)* children in the household? _____

13. During the past year, did . . . have difficulty reading books because they were in English?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

14. How often does . . . read:

a. an English-language newspaper? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

b. magazines in English? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

c. books in English? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

15. How often does . . . read newspapers, magazines, or books in a language other than English? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

16. At any time during the past year, did . . . attend regular school in the U. S.?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

Table V.31 Results of Reverse Scale Analysis for Four Census Questions for Children and Adults by Site, Body of Table Gives the Distribution of Respondents for the Normal and Reversed Scales

Question 6

Scale	Adults				Children			
	San Francisco		Arizona		San Francisco		Arizona	
	Normal	Reversed	Normal	Reversed	Normal	Reversed	Normal	Reversed
Very well	7	1	54	9	3	0	28	3
Well	5	1	28	2	2	3	42	3
Adeq-most	2	3	10	3	0	0	9	0
Adeq-few	0	0	2	4	0	0	9	1
Just a little	1	2	13	0	1	2	10	5
Not at all	0	0	0	0	0	0	2	0
Don't know	0	0	1	0	0	1	0	0
Total	15	7	108	18	6	6	100	12
X^2 Test	N.S.		Sig(.01 level)		N.S.		N.S.	

Question 7

	Adults				Children			
Very well	9	4	54	7	3	0	33	3
Well	3	0	33	6	2	1	41	4
Adeq-most	3	1	4	3	0	1	12	0
Adeq-few	0	1	8	1	0	2	5	1
Just a little	0	1	9	1	1	1	8	4
Not at all	0	0	0	0	0	0	0	0
Don't know	0	0	0	0	0	1	1	0
Total	15	7	108	18	6	6	100	12
X^2 Test	N.S.		N.S.		N.S.		N.S.	

Table V.31 Continued

Question 10

Scale	Adults				Children			
	San Francisco		Arizona		San Francisco		Arizona	
	Normal	Reversed	Normal	Reversed	Normal	Reversed	Normal	Reversed
Very well	8	5	69	8	0	4	40	3
Well	2	3	23	4	1	1	23	4
Adeq-most	0	0	6	2	1	1	6	3
Adeq-few	2	0	5	1	0	0	7	1
Just a little	2	0	8	2	3	0	17	1
Don't know	0	0	1	0	0	0	0	0
Total	14	8	112	17	5	6	93	12
χ^2 Test	N.S.		N.S.		N.S.		N.S.	

Question 11

	Adults				Children			
Very well	8	6	79	8	1	3	41	5
Well	2	2	19	5	1	0	29	5
Adeq-most	0	0	6	1	1	2	9	0
Adeq-few	3	0	1	1	1	1	5	2
Just a little	1	0	7	2	2	0	9	0
Don't know	0	0	0	0	0	0	1	0
Total	14	8	112	17	6	6	94	12
χ^2 Test	N.S.		N.S.		N.S.		N.S.	

N.S. = not significant

H. Monolingual - Bilingual Interviewer Comparison

As discussed previously, a substudy was conducted by RTI to compare the effect of bilingual versus monolingual interviewers. In particular, five pairs of interviewers were assigned to work in five separate areas within each site. Each pair consisted of one monolingual and one bilingual interviewer. The interviewers were randomly selected to participate in the substudy and the sample cases assigned to each pair member were randomized within each area.

In Chapter IV, production results (e.g., average hours per interview) for the two types of interviewers indicated that the monolingual interviewers do as well as or slightly better overall than do the bilingual interviewers. In this section the results of comparing the two types of interviewers by scores obtained from various Census Questions and CP score are investigated. In particular, RTI computed the sample means for 10 CQs and CP score by site and by monolingual and bilingual interviewers. The data used was for children and only for the paired monolingual-bilingual interviewers. The 10 CQs investigated were those examined in Section E as potentially important in predicting a MELP. In addition, to computing the sample means, RTI performed t-tests on these sample means to determine if the hypothesis could be rejected that "the means obtained by monolingual interviewers are not significantly different from those obtained by bilingual interviewers for various CQs and CP score."

The results of these computations are given in Table V.32. The table indicates in general that the sample means for the two types of interviewers are quite similar for the 10 CQs and CP score. Furthermore, the t-tests on the two types of means indicate almost no significant results;

i.e., out of the 55 tests of significance performed at the .05 level only 3 tests (=5.5%) are significant. Thus, from the RTI analysis, there is no evidence that the means obtained on CQs and CP score are different for monolingual and bilingual interviewers. It should also be noted here that RTI also computed crosstabs of interviewers type by response distribution by site for each Census Question. These crosstabs were not analyzed by RTI but were given to CAL for analysis.

Table V.32 Sample Means and Summary of t-tests on Monolingual Versus Bilingual Interviewer Means for Various CQs and CP Score, Data for Children for Paired Interviewers Only [1/]

Variable	Interviewer Type	Site 1	Site 3	Site 5	Site 6	Over Sites
When	Mono	1.90	2.75	2.98	2.24	2.49
	Biling	1.76 ns	2.85 ns	3.00 ns	2.58 ns	2.50 ns
Speak	Mono	3.38	3.22	3.78	3.40	3.46
	Biling	3.23 ns	3.76 *	3.66 ns	3.65 ns	3.60 ns
Und	Mono	3.66	3.45	4.00	3.72	3.71
	Biling	3.43 ns	3.69 ns	3.89 ns	3.69 ns	3.68 ns
Sib	Mono	1.65	1.95	1.78	2.04	1.83
	Biling	1.65 ns	1.76 ns	1.86 ns	2.15 ns	1.84 ns
Frnd	Mono	2.16	2.23	2.13	2.56	2.23
	Biling	2.02 ns	1.87 *	2.23 ns	2.46 ns	2.14 ns
Hlang	Mono	1.03	1.77	1.83	1.52	1.53
	Biling	1.00 ns	1.75 ns	1.77 ns	1.46 ns	1.51 ns
Years	Mono	2.47	1.70	3.34	2.84	2.58
	Biling	2.22 ns	1.85 ns	3.93 ns	2.35 ns	2.54 ns
Birth	Mono	65.9	67.5	65.6	67.2	66.4
	Biling	65.9 ns	67.5 ns	65.2 ns	67.0 ns	66.4 ns
Grade	Mono	5.03	3.77	5.27	4.16	4.65
	Biling	4.90 ns	3.02 ns	5.30 ns	3.85 ns	4.22 *
Ped	Mono	2.85	2.87	2.89	3.68	2.97
	Biling	2.71 ns	3.00 ns	2.75 ns	3.73 ns	3.08 ns
CP	Mono	44.1	41.9	50.2	46.6	45.6
	Biling	41.6 ns	39.3 ns	50.1 ns	49.8 ns	44.4 ns
Sample Size	Mono	68	60	64	25	220
	Biling	51	55	44	26	186

[1/] Data only for the 5 pairs of interviewers (one monolingual and one bilingual) in each site who worked in specially-assigned areas (note no comparison was available in Site 4, S.F. non-Chinese, since all interviewers in San Francisco were Chinese). See Table V.12 for definitions of the various variables.

* = t-Test significant at .05 level.

ns = t-Test not significant at .05 level.

REFERENCES

[1] Cochran, W. G., and Hopkins, C. E., 1961, "Some Classification Problems with Multivariate Qualitative Data", Biometrics, 17, pp. 10-32.

[2] Kendall, M. G., and A. Stuart, 1967, The Advanced Theory of Statistics, Volume 2, Inference and Relationship, Second Edition, Hafner Publ. Co., N. Y., 690 pp.

[3] Lachenbruch, P. A., and Mickey, M. R., 1968, "Estimation of Error Rates in Discriminant Analysis", Technometrics, 10, pp. 1-11.

[4] Kshirsagar, A. M., 1972, Multivariate Analysis, Marcel Dekker, Inc. New York, pp. 187-243.

[5] Nie, N. H. et al., 1975, SPSS Statistical Package for the Social Sciences, Second Edition, McGraw Hill, 675 pp.

[6] Tenenbein, A., 1970, "A Double Sampling Scheme for Estimating from Binomial Data With Misclassifications", Journal of the American Statistical Association, 65, pp. 1351-1361.

[7] Van de Geer, J. P., 1971, Introduction to Multivariate Analysis for the Social Sciences, W. H. Freeman and Company, pp. 243-256.

APPENDIX A

Bilingual Study Census Questionnaire

BILINGUAL STUDY
CENSUS QUESTIONNAIRE

O.M.B. 51-S75048
Expires October 1975

ID No. of DR _____ Sex _____
FI _____ FI No. _____ Date ____
Type (✓): ☐ Self Report ☐ Second Hand Report

1. What is . . .'s date of birth?

   Month _____ Day _____ Year _____

2. What is . . .'s origin or descent? *(USE FLASH CARD A)*

   _____

3. In what state, U. S. territory, or foreign country was . . . born? *(USE FLASH CARD B)*

   _____

   (IF ANSWER IS "this state" OR "different state" SKIP TO Q.5.)

4. When did . . . come to the U. S. mainland to stay?

   _____ 1. 1975
   _____ 2. 1973-1974
   _____ 3. 1971-1972
   _____ 4. 1966-1970
   _____ 5. 1961-1965
   _____ 6. Before 1961
   _____ 7. Don't know

5. Does . . . speak or understand <u>any</u> English?

   _____ 1. Yes
   _____ 2. No *(SKIP TO Q.8)*
   _____ 3. Don't know *(SKIP TO Q.8)*

6. How well does . . . speak English? *(READ ANSWER CHOICES 1-5)*

   _____ 1. Very well
   _____ 2. Well
   _____ 3a. Adequately for most purposes
   _____ 3b. Adequately only for a few purposes
   _____ 4. Just a little
   _____ 5. Not at all
   _____ 6. Don't know

7. How well does . . . understand spoken English? *(READ ANSWER CHOICES 1-5)*

   _____ 1. Very well
   _____ 2. Well
   _____ 3a. Adequately for most purposes
   _____ 3b. Adequately only for a few purposes
   _____ 4. Just a little
   _____ 5. Not at all
   _____ 6. Don't know

8. What *(OTHER)* languages does . . . speak? *(USE FLASH CARD C)*

   _____

   (IF NONE, SKIP TO Q.12. IF ONLY ONE, SKIP TO Q.10)

9. Which one of these languages does . . . speak most often? *(USE FLASH CARD C)*

10. How well does . . . speak *(PRINCIPAL LANGUAGE FROM Q.8 OR Q.9)?*
(READ ANSWER CHOICES 1-4)

_____ 1. Very well
_____ 2. Well
_____ 3a. Adequately for most purposes
_____ 3b. Adequately only for a few purposes
_____ 4. Just a little
_____ 5. Don't know

11. How well does . . . understand *(PRINCIPAL LANGUAGE FROM Q.8 OR Q.9)?*
(READ ANSWER CHOICES 1-4)

_____ 1. Very well
_____ 2. Well
_____ 3a. Adequately for most purposes
_____ 3b. Adequately only for a few purposes
_____ 4. Just a little
_____ 5. Don't know

12. What language does . . . usually speak when talking to: *(USE FLASH CARD C)*

a. brothers and sisters? _____
b. parents? _____
c. other older relatives? _____
d. . . .'s best friend? _____
e. *(IF . . . IS AN ADULT)* children in the household? _____

13. During the past year, did . . . have difficulty reading books because they were in English?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

14. How often does . . . read:

a. an English-language newspaper? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

b. magazines in English? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

c. books in English? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

15. How often does . . . read newspapers, magazines, or books in a language other than English? *(READ ANSWER CHOICES)*

_____ 1. Often
_____ 2. Occasionally
_____ 3. Not at all

16. At any time during the past year, did . . . attend regular school in the U. S.?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

17. During the past year, did . . . take any courses at business, vocational or technical school?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

(IF "NO" OR "DON"T KNOW" TO BOTH Q's 16 AND 17, SKIP TO Q.20)

18. In any school or course attended during the past year, was . . . taught in a language other than English?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

19. During the past year has a teacher, counselor, or school official said that . . . had difficulty speaking or understanding English?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

20. At any time during the past year did . . . take any course or class for people whose principal language is not English?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

21. What is the highest grade or year of regular school . . . has ever attended? *(USE FLASH CARD D)*

_____

(IF "NONE" SKIP TO 27. IF "DON'T KNOW," SKIP TO Q.23)

22. How many years of . . .'s schooling was taught in English? _____

23. Did . . . speak English before going to school for the very first time?

_____ 1. Yes
_____ 2. No *(SKIP TO Q.25)*
_____ 3. Don't know *(SKIP TO Q.25)*

24. How well did . . . speak English before going to school for the very first time? *(READ ANSWER CHOICES 1-4)*

_____ 1. Very well
_____ 2. Well
_____ 3. Adequately
_____ 4. Just a little
_____ 5. Don't know

25. Has . . . ever repeated a grade in school?

_____ 1. Yes
_____ 2. No *(SKIP TO Q.27)*
_____ 3. Don't know *(SKIP TO Q.27)*

26. What grade(s) did . . . repeat? _____

27. Does . . . have any difficulty in speaking or understanding English? *(READ ANSWER CHOICES 1-5)*

_____ 1. Yes, difficulty in both speaking and understanding
_____ 2. Yes, difficulty only in speaking
_____ 3. Yes, difficulty only in understanding
_____ 4. Yes, doesn't speak or understand at all
_____ 5. No, no difficulty in speaking or understanding
_____ 6. Don't know

28. Does . . . prefer to avoid places where only English is spoken?

_____ 1. Yes
_____ 2. No
_____ 3. Don't know

29. During the past year has . . . been employed at any time?

_____ 1. Yes
_____ 2. No *(SKIP TO Q.31)*
_____ 3. Don't know *(SKIP TO Q.31)*

30A. For whom did . . . work? *(NAME OF COMPANY, BUSINESS, ORGANIZATION, OR OTHER EMPLOYER)*

_____

30B. What kind of business or industry is this? *(FOR EXAMPLE, TV AND RADIO MANUFACTURING, RETAIL SHOE STORE, STATE LABOR DEPARTMENT, FARM)*

_____

30C. What kind of work did . . . do? *(FOR EXAMPLE, ELECTRICAL ENGINEER, STOCK CLERK, TYPIST, FARMER.)*

_____

30D. What were . . .'s most important activities or duties? *(FOR EXAMPLE, TYPES, KEEPS ACCOUNT BOOKS, FILES, SELLS CARS, OPERATES PRINTING PRESS, FINISHES CONCRETE)*

_____

30E. At work, what language does . . . usually speak? *(USE FLASH CARD C)*

_____

31. What is the usual language spoken in this household? *(USE FLASH CARD C)*

_____

32. What other languages are spoken in this household? *(USE FLASH CARD C)*

_____

APPENDIX B

Additional Results of Discriminate Analysis Using FCTR and CP to Define LESA

Table B-1 Results of Using Discriminant Analysis to Predict Whether A Child is LESA or Non-LESA, Discriminant Function Based on 10 Census Questions, LESA Group Defined by FCTR[1/] [2/]

	FCTR[4/] Used to Define LESA									
	Site 1 (Miami)		Site 3 (El Paso)		Site 4 (S.F. Non-Chinese)		Site 5 (N.E. Arizona)		Site 6 (S.F. Chinese)	
FCTR Group	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA
LESA	121	34	171	25	16	6	59	20	47	11
Non-LESA	34	128	29	139	18	93	45	136	20	68
α_1[3/]	21.9		12.8		27.3		25.3		19.0	
α_2	21.0		17.3		16.2		24.9		22.7	
% Correct	78.6		85.2		82.0		75.0		78.8	
Est % LESA	48.9		54.9		25.6		40.0		45.9	
Actual % LESA	48.9		53.8		16.5		30.4		39.7	

Over Sites

FCTR Group	Predicted Group LESA	Non-LESA
LESA	398	112
Non-LESA	141	569
α_1	22.0	
α_2	19.9	
% Correct	79.3	
Est % LESA	44.2	
Actual % LESA	41.8	

1/ 10 Census Questions = When, Speak, Und, Sib, Frnd, Hlang, Years, Birth, Grade and Ped.

2/ Discriminant Function Computed for each stie.

3/ α_1 = Percent of LESAs classified as Non-LESA; α_2 = Percent of Non-LESAs classified as LESA; % correct = overall percent classified correctly.
Est % LESA = estimated percent of LESAs = 100 x (Number predicted as LESA)/(Total number of children)
Actual % LESA = percent of LESAs as given by FCTR.

4/ LESA defined as FCTR ≤.2.
Note Site 5 includes both Ganado and Window Rock.

Table B-2 Results of Using Discriminant Analysis to Predict Whether A Child is LESA or Non-LESA, Discriminant Function Based on Spund, Usage and Years, LESA Defined by FCTR and CP[1/]

FCTR[3/] Used to Define LESA

	Site 1 (Miami)		Site 3 (El Paso)		Site 4 (S.F. Non-Chinese)		Site 5 (N.E. Arizona)		Site 6 (S.F. Chinese)	
FCTR Group	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA	Predicted Group LESA	Non-LESA
LESA	122	33	160	36	17	5	59	20	40	18
Non-LESA	41	121	28	140	25	86	46	135	23	65
α_1[2/]	21.3		18.4		22.7		25.3		31.0	
α_2	25.3		16.7		22.5		25.4		26.1	
% Correct	76.7		82.4		77.4		74.6		71.9	
Est LESA	51.4		51.6		31.6		40.4		43.2	
Actual LESA	48.9		53.8		16.5		30.4		39.7	

Over Sites

FCTR Group	Predicted Group LESA	Non-LESA
LESA	391	119
Non-LESA	147	563
α_1	23.3	
α_2	20.7	
% Correct	78.2	
Est LESA	44.1	
Actual LESA	41.8	

Table B-2 Continued

CP Used to Define LESA

CP Group	Site 1 (Miami) Predicted Group LESA	Site 1 (Miami) Predicted Group Non-LESA	Site 3 (El Paso) Predicted Group LESA	Site 3 (El Paso) Predicted Group Non-LESA	Site 4 (S.F. Non-Chinese) Predicted Group LESA	Site 4 (S.F. Non-Chinese) Predicted Group Non-LESA	Site 5 (N.E. Arizona) Predicted Group LESA	Site 5 (N.E. Arizona) Predicted Group Non-LESA	Site 6 (S.F. Chinese) Predicted Group LESA	Site 6 (S.F. Chinese) Predicted Group Non-LESA
LESA	129	31	170	63	26	8	69	23	54	20
Non-LESA	36	121	18	113	29	70	41	127	20	52
α_1	19.4		27.0		23.5		25.0		27.0	
α_2	22.9		13.7		29.3		24.4		27.8	
% Correct	78.9		77.8		72.2		75.4		72.6	
Est LESA	52.1		51.6		41.4		42.3		50.7	
Actual LESA	50.5		64.0		25.6		35.4		50.7	

Over Sites

CP Group	Predicted Group LESA	Predicted Group Non-LESA
LESA	442	151
Non-LESA	135	492
α_1	25.5	
α_2	21.5	
% Correct	76.6	
Est LESA	47.3	
Actual LESA	48.6	

1/ Discriminant function computed for each site.

2/ Definitions of α_1, α_2, % correct, Est % LESA and actual % LESA given in Table B-1.

3/ FCTR ≤.2 = LESA
CP ≤50 = LESA

Note Site 5 includes both Ganado and Window Rock.